The American Economic Review

ARTICLES

The 1974 Report of the President's Council of Economic Advisers:

A. O. KRUEGER — International Issues

D. LAIDLER — The Control of Inflation and the Future of the International Monetary System

F. MODIGLIANI — A Critique of Past and Prospective Economic Policies

W. D. NORDHAUS — Energy in the Economic Report

P. A. SAMUELSON AND S. SWAMY — Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis

R. N. BATRA — Technological Change in the Soviet Collective Farm

T. JOHNSON AND F. J. HEBEIN — Investment in Human Capital and Growth in Personal Income 1956–1966

J. A. HANSON — Optimal International Borrowing and Lending

A. TAKAYAMA — On Biased Technological Progress

J. V. HENDERSON — The Sizes and Types of Cities

G. W. DOUGLAS AND J. C. MILLER III — Quality Competition, Industry Equilibrium, and Efficiency in the Price-Constrained Airline Market

S. ROSEFIELDE — Factor Proportions and Economic Rationality in Soviet International Trade 1955–1968

R. E. SCHULER — The Interaction Between Local Government and Urban Residential Location

SHORTER PAPERS: A. Razin; A Sandmo; E. K. Browning; A. P. Lerner; J. L. Stein; R. Fels and C. E. Hinshaw; H. O. Stekler; W. G. Waters II; E. Silberberg; S. S. Sidhu; H. M. Hochman and J. D. Rodgers; J. E. Anderson; W. Schulze and R. C. d'Arge; J. R. Marsden, D. E. Pingry, and A. Whinston; J. M. Griffin; M. L. Greenhut, M. J. Hwang, and H. Ohta; J. E. Olson; W. Ramm; M. B. Krauss; D. Salkever; J. M. Parkman and S. Y. Wu; H. Hori; E. F. Fama and A. B. Laffer.

SEPTEMBER 1974

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

• Published at George Banta Co., Inc., Menasha, Wisconsin.

• THE AMERICAN ECONOMIC REVIEW, including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory*, and Supplements, is published by the American Economic Association and is sent to all members six times a year, in March, May, June, September, October, and December.

• Membership dues of the Association are $23.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature*. Junior membership at $11.50 per year is available to registered students for a maximum of three years. Subscriptions (libraries, institutions, or firms) are $34.50 a year, and only subscriptions to both publications will be accepted. Single copies of the *Review* and *Journal* are $5.00 each plus a $.50 per order service charge. Orders should be sent to the Secretary's office, Nashville, Tennessee.

• Correspondence relating to the *Papers and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, Rendigs Fels, 1313 21st Avenue South, Nashville, Tennessee 37212. Change of address must reach the Secretary by the 1st of the month prior to the month of publication. The Association's publications are mailed second class and are not forwarded by the Post Office.

• Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

THE AMERICAN ECONOMIC REVIEW

September 1974

VOLUME LXIV, NUMBER 4

P 35738

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A $10 submission fee must accompany each manuscript. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

Articles

The 1974 Report of the President's Council of Economic Advisers: International Issues *Anne O. Krueger* 523

The 1974 Report of the President's Council of Economic Advisers: The Control of Inflation and the Future of the International Monetary System *David Laidler* 535

The 1974 Report of the President's Council of Economic Advisers: A Critique of Past and Prospective Policies *Franco Modigliani* 544

The 1974 Report of the President's Council of Economic Advisers: Energy in the Economic Report *William D. Nordhaus* 558

Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis *P. A. Samuelson and S. Swamy* 566

Technological Change in the Soviet Collective Farm *Raveendra N. Batra* 594

Investment in Human Capital and Growth in Personal Income 1956–1966 *Thomas Johnson and Frederick J. Hebein* 604

Optimal International Borrowing and Lending *James A. Hanson* 616

On Biased Technological Progress *Akira Takayama* 631

The Sizes and Types of Cities *J. V. Henderson* 640

Quality Competition, Industry Equilibrium, and Efficiency in the Price-Constrained Airline Market *George W. Douglas and James C. Miller III* 657

Factor Proportions and Economic Rationality in Soviet International Trade 1955–1968 *Steven Rosefielde* 670

The Interaction Between Local Government and Urban Residential Location *Richard E. Schuler* 682

Shorter Papers

A Note on the Elasticity of Derived Demand Under Decreasing Returns *Assaf Razin* 697
A Note on the Structure of Optimal Taxation *Agnar Sandmo* 701
The Diagrammatic Analysis of Multiple Consumption Externalities *Edgar K. Browning* 707
Priorities and Pollution:
Comment *Abba P. Lerner* 715
Reply *Jerome L. Stein* 718
An Analysis of Turning Point Forecasts:
A Fairly Polite Comment *Rendigs Fels and C. Elton Hinshaw* 724
A Polite Reply *H. O. Stekler* 728
Transport Costs and the Static Welfare Costs of Tariffs *W. G. Waters II* 730
The Theory of the Firm in "Long-Run" Equilibrium *Eugene Silberberg* 734
Relative Efficiency in Wheat Production in the Indian Punjab *Surjit S. Sidhu* 742
Redistribution and the Pareto Criterion *Harold M. Hochman and James D. Rodgers* 752
A Note on Welfare Surpluses and Gains From Trade in General Equilibrium *James E. Anderson* 758
The Coase Proposition, Information Constraints, and Long-Run Equilibrium *William Schulze and Ralph C. d'Arge* 763
The Process Analysis Alternative to Statistical Cost Functions:
Comment *J. R. Marsden, D. E. Pingry, and A. Whinston* 773
Reply *James M. Griffin* 777
Price Discrimination by Regulated Motor Carriers:
Comment *M. L. Greenhut, M. J. Hwang, and H. Ohta* 780
Reply *Josephine E. Olson* 785
On the Durability of Capital Goods Under Imperfect Market Conditions *Wolfhard Ramm* 787
Commodity Trade and Factor Mobility *Melvyn B. Krauss* 797
Choice Involving Unwanted Risky Events and Optimal Insurance:
Comment *David Salkever* 802
Reply *J. M. Parkin and S. Y. Wu* 804
The Number of Firms and Competition:
Comment *Hajime Hori* 805
Reply *Eugene F. Fama and Arthur B. Laffer* 806

Notes 807

The 1974 Report of the President's Council of Economic Advisers: International Issues

By ANNE O. KRUEGER*

The international coverage in the 1974 Economic Report is not comparable with that of previous years. In the past, the international chapter was devoted primarily to the *U.S.* balance of payments and to policies designed to improve it. The year 1973 saw the abandonment of fixed exchange rates. Consequently less than three pages are devoted to discussion of balance-of-payments figures (and none to policies to "remedy" them), mostly explaining why balance-of-payments data no longer have their former interpretations. Indeed, the official transactions balance now " . . . approximately shows the extent to which governments have bought or sold currencies to influence the exchange rate . . . " (p. 193).

This year's chapter is instead addressed to a description of developments, primarily in the international monetary arena, over the past year, and of American policy stances toward international monetary negotiations and international trade. By and large, the effects of international events on the American economy are dealt with in earlier chapters of the Report.

I. International Monetary System

The year 1973 was a truly amazing one for the international economy. The international monetary system was *de facto* transformed to one of floating rates—with official intervention—by March 19, 1973. In a year marked by economic difficulties, the most spectacular economic success was a nonevent—the absence of major foreign-exchange market disruption—in the face of the Middle East oil producers' attempt to exercise their joint monopoly power late in the year.

The Council treats the year's events chronologically. The first quarter saw the last massive speculative capital flows and the abandonment of fixed exchange rates; in the second quarter, the dollar depreciated substantially; in the third quarter, there was " . . . limited intervention by a number of central banks, including the United States, to prevent the dollar from rising . . . " (p. 183). The fourth quarter witnessed appreciation of the dollar with "substantial" intervention by central banks to slow the rise.

A. *Abandonment of Fixed Rates*

The Council's interpretation of the abandonment of fixed rates is somewhat superficial:

> Demand pressures were intensified when large speculative capital flows led to excessive increases in the monetary supply in a number of countries. [p. 181] . . . [M]assive capital flows from the United States to Europe and Japan had two effects: First, foreign central banks added around $10 billion in claims against the United States to their reserves as a result of their efforts to support the value of the dollar in terms of

* Professor of economics, University of Minnesota. This article was written while I was visiting professor of economics at the Massachusetts Institute of Technology. I am indebted to Richard Eckaus, Charles Kindleberger, and Rachel McCullough for helpful comments on an early draft of this paper, and to Peter Kenen for stimulating discussion of some of the issues raised here.

> their own currencies. Second, when large-scale market intervention failed to restore stability to foreign exchange markets, fixed exchange rates were abandoned; . . . [p. 183]

Here and elsewhere in the chapter, there is a curious tendency to describe events, rather than analyze them.[1] In the above quoted instance, the speculative capital flows seem totally exogenous. Yet, in fact, speculative capital flows and world inflationary pressures were both partly the consequence of huge prior accumulation of dollar assets by foreigners, particularly in 1971. That accumulation, in turn, had its origins in the overvaluation of the dollar and other features of fixed exchange rates under the dollar standard.

Throughout the Report, world inflationary pressures of 1973 are treated as if they were totally independent of past American behavior. This is true not only in the international chapter, but also in chapters on domestic economic matters when sources of inflation in the United States are discussed. The facts are that foreign holdings of foreign exchange rose from $32 billion at the end of 1969 to $78 billion at the end of 1971, $103 billion at the end of 1972, and $117 billion in March 1973—with total reserves of the rest of the world increasing from $78 billion in December 1969 to $179 billion by March 1973. Of that $101 billion increase, $55 billion was in *U.S.* liabilities, and the American payments deficit must be assigned primary responsibility for the increase in world liquidity. One need not be a monetarist to believe such a massive increase was bound to have inflationary consequences.[2] Indeed, Michael Parkin warned of the problem in his review last year.

A significant portion of the chapter is given to discussion of exchange rate realignments in 1973 and a supplement to the chapter discusses "Measurement of Effective Changes in Exchange Rates." There is a useful table (53, p. 197) summarizing the percentage changes in exchange rates of various countries during the year by type of exchange-rate regime prevailing at the end of the year (joint float, individual float, and intervention to maintain dollar, pound, franc, or average parity).

The supplement is essentially a cookbook guide as to how "effective changes in exchange rates" are computed. The differences between using export shares, import shares, and trade shares in weighting changes in different currencies are explained, as is the difference between using dollar prices of foreign currencies and foreign prices of the dollar ("If the value of the mark expressed in dollars rises by 33 percent, for example, then the value of the dollar expressed in marks falls by only 25 percent . . . " p. 220).

No reason is given for wanting a measure of effective changes other than that it is "useful." Issues such as why gold or *SDR*s are not a numeraire of the system and why the trade shares rather than current account or transactions shares in payments and/or receipts are not the more appropriate weights for the index are not touched upon. However, the Council does alert readers to the fact that more than one index is possible: it presents five and explains the differences between them without choosing among them. The five include Morgan Guaranty Trust's 14-country index, Reuters Currency Index for 9 countries, one from the Treasury (22 countries), and one from the Council (weighting exchange rate changes by shares in world trade to reflect the fact that " . . . changes

[1] In this regard, the Council's use of "stability" as synonomous with the absence of private speculation under fixed rates is unprofessional. The term is used, with the same meaning, again on page 184.

[2] All that needs to be accepted is that some countries, which otherwise would have adopted restrictive policies or devalued, did not do so because of greater liquidity.

in any one exchange rate affect trade of other countries as well . . . " p. 224). The fifth index is derived from a Central Intelligence Agency (sic) trade flow model, whose properties cannot be inferred from the one-paragraph description given of it. The model evidently is used to estimate the sensitivity of various countries' commodities to price changes, and to alter weights of different exchange rates accordingly.

B. *Managed Floating*

By March 1973, the major currencies were all floating vis-à-vis each other. In view of the shocks withstood by the system late in the year, the Council's judgment is restrained:

> . . . Early in the year the governments of most major countries abandoned attempts to fix exchange rates at negotiated levels. While central banks continued to intervene to some extent, foreign exchange markets played the major role in determining the exchange rates that would clear the market. This process was marked at times by unusually large fluctuations of market exchange rates. Nevertheless, the market performed its intermediating function well, and neither trade nor long-term capital flows were seriously disrupted at any time during the year. . . . [p. 182]

There is, unfortunately, no discussion or statement of the criteria by which the Council judges exchange rate changes to have been "unusually large." The float started after years of market intervention had left a huge "dollar overhang." In that circumstance, it was reasonable to suppose that the stock equilibrium rate and the flow equilibrium rate diverged, and perhaps even widely. There was also the consideration that even ardent advocates of flexible exchange rates had conceded that transition to a float might cause initial disturbances. In fact, the dollar depreciated approximately 11 percent (Morgan Guaranty's Index) from February to July, and then appreciated by about 6 percent between July and December. In light of the dollar overhang and the state of disequilibrium at the time the float was inaugurated, it would seem perhaps more convincing to conclude that the inauguration of floating rates was surprisingly smooth.

For evaluating present and future economic policy, it probably doesn't make much difference whether one thinks that 1973 exchange rate fluctuations were "large" or "small." What does make a difference is that the float emerged "dirty." As described by the Council:

> . . . While governments have continued to intervene in the foreign exchange market in order to influence the movement of the exchange rate, it can no longer be assumed that they will finance an excess demand or absorb an excess supply of foreign currencies at given exchange rates. The situation can best be described as one of managed floating. [p. 196]

Evaluation and analysis of the managed float is intertwined with two related issues: the impact of the oil situation on the American international position and on the world; and the future of the international monetary system.

Turning to the oil situation, the first and obvious reaction of any economist to the authors of the Report must be one of sympathy and commiseration. January 1974 was about as difficult a time as there could have been to write the Report. To refresh the reader's memory, it will be recalled that the increased price of Middle East crude oil led to substantial *ex ante* current account deficits on the part of most Western European countries, and responsible *conservative* estimates put the increase in Middle East oil producers' current account receipts—the OECD's current account deficit—at between $30 and $50 billion in 1974.

While one could argue about which countries might spend how much of their

increased revenues on goods and services, and whether the $30–50 billion was a one-shot affair or would be likely to continue for several years (and, for that matter, how long the cartel could stay together, what the intermediate-run elasticity of demand was, and whether or not the producers had overshot their short-run and/or long-run monopoly price), there seemed to be little doubt that Western Europe, Japan, and North America would collectively run a current account deficit of that order of magnitude in 1974. Particularly in December and January, forecasts of a considerably more pessimistic nature were rampant.

That a deficit of that amount could be financed did not seem to be a major problem: the oil exporting countries could accumulate dollar-denominated assets in the United States or the Euro-dollar market, or European claims, although it seemed likely that dollar-denominated claims would be preferred. Only an extreme optimist would have dared to hope that the existing dollar overhang would be transferred by its holders to the oil exporters. Such a solution would have been ideal from the viewpoint of the United States and would have entailed no American adjustment cost other than the fact of deteriorated terms of trade (which estimates implied would be $13 billion in oil import bill, or less than 1 percent of *GNP* and one-third of one year's real growth) and the real income loss associated with it.

The real question, and the one whose answer was perhaps least clear in January, was the degree to which Europe and Japan would attempt to maintain their current accounts by depreciating their exchange rates, pushing exports, and reducing imports. Such an effort, while destined to be at best partially successful, would have had frightening implications for the international monetary system, as it would have placed the entire burden of adjustment on current account with the United States, thus auguring a current account shift of virtually incomprehensible size from 1973 to 1974. To compound fears, the French chose mid-January as the time to abandon their participation in the joint European float, thus giving tangible support to the view that the United States might have to bear much of the real adjustment.

One could, then, as of January, imagine all sorts of nightmarish scenarios: pressures for the Burke-Hartke bill become overwhelming as the United States is flooded with cheap imports; the United States intervenes in the foreign exchange market, buying foreign exchange, as European Central Banks are intervening and buying dollars. With the French decision to float, these possibilities appeared all too real. Moreover, concern about them implied the need for international economic cooperation and diplomacy once again. Under these circumstances, one cannot ask for an official American document in which these concerns are stated baldly. The consequence is a very muted discussion of the "... fairly harmonious ..." (p. 199) relations among countries, the need for "... conventions ..." to regulate intervention (p. 200) with only one mention of the possibility of "... unilateral measures at the expense of one's neighbors ..." by "... just one nation, and then others ... the resulting disruption of international trade ..." (p. 220). Only once is the fear of "... a round of competitive deflation, depreciation, or trade restriction ..." (p. 35) explicitly mentioned.[3]

The above-listed considerations made it extremely difficult to write the Report, both because of uncertainty, and for diplomatic reasons. The possibilities also quite clearly demonstrate the problem with

[3] See also the statement of Secretary Schultz to the Washington Energy Conference on February 11, 1974.

"managed floats," be that management through exchange-market intervention or through capital controls: when countries have current account targets, they can control capital flows and intervene in the foreign exchange market to achieve their targets.[4] There is still the problem of reconciling conflicting *ex ante* objectives with respect to the size of the current account. There must either be an *n*th country (the United States?) whose current account is equal to the negative of the sum of other countries' current accounts, or there must be a mechanism for reconciling inconsistent aspirations *ex ante*. Such a mechanism is "international consultation and cooperation."

The nice, neat "floating rate" of the classroom is not to be and cannot be. Richard Cooper's insights are, in that regard, more relevant than those derived from the model of a flexible exchange rate working in a free market. International economic interdependence is too great, and the United States cannot or will not any longer play a passive role to let others achieve their objectives. Negotiations among countries will have to do the job, and managed floating implies at least as large an element of political decision making as did attempts to maintain fixed, misvalued exchange rates. Whether mutually conflicting current account targets will be reconciled by finance ministers and central bankers, or by foreign ministers, remains to be seen. The alternative to such behind-the-scenes negotiations is, however, probably the frightening spectre of trade warfare.[5]

It is in this context that the Council's evaluation that 1973 exchange rate fluctuations were "excessive" becomes important. In fact the turning point in the dollar's fortunes came in July *after* it was announced that the Treasury would intervene in the market. While actual intervention was very small, the announcement itself certainly helped reverse expectations.[6]

In a sense, it is as important to derive "criteria" for intervention under floating rates as "objective indicators" were under the Schultz proposals.[7] The international chapter's section on the future of the international monetary system, unfortunately, reiterates the Shultz proposals. This is not the place to review those proposals, which were essentially an attempt to simulate floating rates (with discrete but hopefully small changes in rates at frequent but uncertain intervals) on the basis of "presumptive rules of adjustment."

Even before the events of February and March 1973, it was doubtful whether any sort of agreement to reform the system along the lines of those proposals would be negotiated. Quite aside from concern over speculation which might result from the presumptive rules, and over the fact that current account objectives were not to be reconciled, there did not seem to be sufficient common interests to induce countries to agree quickly to the Shultz—or any other—proposal. After March, pressures for agreement were, if anything, di-

[4] Hence the detailed table, pp. 198–99, listing the changes in capital controls in 1973, country by country. See Section Ic. below.

[5] A shift in the current account of the United States of some $40 billion (since 1973 was a surplus year) would entail massive short-term resource reallocational requirements which involve costs. Even if there were no recession resulting from such shifts, it would be inefficient to get such massive real resource changes to accommodate a two- or three-year OECD deficit.

[6] Because of the overhang, there was considerable reason to expect that when the dollar did begin appreciating, it might appreciate rapidly as expectations reversed. In fact, appreciation was relatively slow until the announcement of the oil embargo.

[7] There is a difference, and that difference may in fact be important: under fixed rates, Central Banks intervened in the market automatically unless there was a major policy decision to the contrary; intervention under floating rates will now require a conscious decision. In that sense, the burden of proof has shifted toward those who wish to intervene in the foreign exchange market.

minished and it seemed evident by January that the initiative for return to fixed rates was a dead letter. Nonetheless, it was still official policy, and not until late February did the Group of Twenty officially abandon the quest for a return to fixed rates. The Council, tied to official policy, could therefore only write of the Shultz proposals in January. Discussion of the future of the managed float is therefore unsatisfactory because of the constraints placed upon the members of the Council by Administration policy.

An initiative to devise means for agreeing on mutually consistent current account targets is sorely needed, and will be difficult to achieve. One wishes that the Council had been able to devote its attention to that inherently difficult question, but it could not do so. By its nature, the problem of reconciling current account targets must at present be a subject of diplomatic negotiations. It seems reasonably clear that some agreement was reached with the French after their January announcement. To date, it would appear that there has been enough consultation so that the risks inherent in allocating the oil exporters' current account surplus among the importers have been satisfactorily handled.

Despite these problems and the importance of their satisfactory resolution for the future of the system, one must, on the whole, give the Administration good grades for its international monetary policy in 1973: the performance of the managed rate system was reasonably good, and a vast improvement over the earlier crisis-followed-by-crisis situation. The managed float is probably a better system than any alternative currently available. In the first year under floating rates, the shock to expectations and the initial impact of oil price increases were both handled without major disturbances despite the fact that the initial position was clearly one of severe disequilibrium. Perhaps, despite intervention, market forces manage to make themselves felt more clearly under the float, and there is a presumption for price adjustments. It may be that that change in itself accounts for improved performance to date.

C. *International Investment*

It was shown above that there are dangers of inconsistent current account targets under the managed float, and that there is a need for international negotiation to reconcile conflicting targets in the absence of well-defined objective criteria for intervention. The probability of competitive devaluation is not high, but the cost of such an outcome would be enormous.

There are two equivalent ways in which a country can seek to attain its current account target under floating rates: its Central Bank can intervene directly in the foreign exchange market or it can so regulate private international capital flows that the net capital account balance is the negative of the desired current account figure. In practice, such regulations could probably not attain targets with any degree of precision, but capital controls could almost certainly achieve zero current account balances.

Under these circumstances, it is small wonder that the international chapter of the Report devotes attention to changes in capital controls in 1973. There is a two-page table (54, pp. 198–99) itemizing changes for the maior OECD countries, although the account of events during 1973 is highly descriptive, and it is left to the reader to infer the significance of the reported moves.

One of the major achievements of American international economic policy in 1973–74 was the removal of the capital controls including the interest equalization tax, restrictions on American corporations' direct investment abroad, and

limitations on bank lending which had been imposed during the years of balance-of-payments problems. Those controls were substantially relaxed in December 1973, and removed entirely in January 1974. The Report mentions this move (p. 200) in one paragraph. The under-emphasis accorded to the removal of controls, relative to the achievement, is surprising.

II. International Trade Policy in 1973

The purpose of maintaining a smoothly functioning international monetary system is to enable the free and orderly exchange of goods and services between countries. Whereas the Administration's international monetary policy must be judged fairly favorably, its performance with respect to international trade issues was less satisfactory.

During 1973, the following were among the actions in the trade arena taken by, or with strong initiatives from, the United States: 1) a new multilateral textile "arrangement" was negotiated which in effect places quantitative restrictions on American textile imports via "voluntary restraint" by exporting countries; 2) an embargo was placed upon soybean exports, which was followed by export controls on soybeans and other high-protein feeds; 3) "Project Independence" was announced to assure domestic oil supplies and insulate the United States from the world oil market; 4) quotas on imports of meat were suspended and those for dairy products were increased "... temporarily ..." (p. 210); 5) controls were imposed on the export of steel scrap and the Japanese agreed to import less scrap and fewer logs (p. 211); and 6) compensation was demanded from the EEC for losses incurred by the enlargement from six to nine members, and tariff retaliation was threatened unless satisfactory compensation was given. The initial demands were unreasonably high in the judgment of most observers.

The textile quota is certainly not the only voluntary restraint negotiated by the United States, quite aside from those on agricultural imports. As Robert Baldwin and the Tariff Commission have documented, there are nontariff barriers of a variety of forms imposed by the United States, and their total effect on trade (both directly and through implied threats to successful entrants to the *U.S.* market) is probably substantial.[8] It can be argued that the nontariff barriers are quantitatively important, not only in their overall restrictive effect on American trade, but also in their particularly heavy incidence on the products of the developing countries. Although the Administration has opposed the Burke-Hartke and related bills, its willingness to adopt voluntary restrictions must be suspect, regardless of its professions of advocacy of free trade.

Similarly, the embargo and subsequent licensing of soybean exports was ludicrous. The United States had devalued twice in the preceeding two years; its stated policy goal was to expand exports and to shift the current account by a net $13 billion. Extreme pressures had been placed on the Japanese, major soybean importers, to appreciate and expand their imports from the United States. Not only did the soybean embargo go against these objectives, it hurt the American posture in efforts to deal with foreign supply disruptions. Ironically, the day of reckoning came soon: the American stance after the Arab oil "embargo" was considerably weakened, not only by past treatment of the Soviet Union, Eastern Europe, China, and Cuba,

[8] As this paper was being completed, the April 22 *New York Times* reported a General Accounting Office estimate placing the cost of the textile agreement in 1972 at $276–$632 million in terms of higher prices to consumers alone.

but more dramatically and more recently by the soybean decision.

Project Independence is equally controversial. It is dealt with in a separate chapter on Energy and Agriculture in the Report and is of concern here only for its international implications. In light of American policy with soybeans, it is amusing to read the statement of principles with respect to trade in oil: "... while the Nation needs to be protected from dependence on unreliable supplies, domestic producers should not be isolated from normal business risks ... " (p. 123).

The suspension and expansion of agricultural quotas is, of course, to be welcomed to the extent that it reflects an underlying reduction in reliance upon quantitative restrictions. To the extent it instead reflects the propensity to manipulate controls in response to short-term domestic needs, it is questionable whether there is any positive benefit to the international trade network.

The tone of the section on international trade relations is, repeatedly, "we all intervene with trade." After mention of the lifting of meat and dairy quotas, the Council reports: "Many other countries took similar actions. Japan, for instance, reduced tariffs on about 5 percent of her import categories and expanded or eliminated some of the remaining quotas on imports of manufactured and agricultural products ... " (p. 210). The account continues with a listing of *tariff* reductions by other countries.

Likewise, the

> ... increase in domestic prices as a result of foreign demand pressure caused considerable resentment in many producing countries and led to public demands that domestic supplies be protected. The situation was further aggravated when some governments imposed price controls in an effort to contain the inflationary pressures. Since export prices remained uncontrolled, domestic producers had an increased incentive to export their goods.... To alleviate the [resulting] domestic shortages, a number of governments imposed controls on exports. [p. 210]

Only then is the reader told that the United States behaved in the above-described way, and no mention is made of the fact that we had devalued our currency in an effort to expand exports while the other culprits had in fact been appreciating at American insistence.

The same "we are all alike" theme pervades the description of the American position on international trade negotiations. The object of these negotiations is "... continued dismantling of trade barriers ... " (p. 213). The Council lists five areas in which negotiations must be undertaken:

> ... nontariff barriers related to domestic economic and social policies; agricultural trade barriers related to domestic agricultural programs; safeguard measures to facilitate an orderly adjustment to new market conditions by producers in *importing* countries; subsidies and other government assistance to industries; and finally new understandings on the access of consuming countries to sources of supply ... [p. 213]

Although each of these five areas is discussed, there is a disappointing absence of concrete proposals for these areas, and the Council appeals for support of the Administration's proposed trade bill, which would mandate negotiating authority for reduction of *tariff* barriers. The bill is a two-edged sword: the President could impose import surcharges or limit imports for balance of payments or other reasons. He could also suspend restrictions, retaliate against other countries imposing import restrictions against the United States, and so on. For "good" purposes, the bill would grant the necessary authority. In the hands of a protectionist or *ad hoc* adminis-

tration, however, it could be a disaster. One's interpretation of the Administration's behavior over the past year—interventionist or free marketeer constrained by political forces beyond its control—is critical in assessing the trade bill. The very strong compensation demands made of the EEC and the soybean experience are not reassuring.

One has the uncomfortable feeling, moreover, that American policy has been somewhat short-sighted. The Administration objected to the view that the American economy was all-powerful, and appears to have gone too far in the direction of regarding the United States as being "just like" any other nation. Perhaps the most telling case in point was its failure to do some of the obvious things that were within its power. The soybean export embargo is the best illustration. Even if it was so that something had to be done,[9] surely there could have been consultations with Japan. Simple acknowledgement that there were contracts for delivery and that the position was such that the United States was forced to back down on its commitments would have been superior to unilateral abrogation of contracts.

Better yet, and probably sorely needed in the supply-short inflationary world that seems to be with us, would be an American initiative for a cooperative international agency which could act in situations of sharp, temporary shifts in supply. It is surely in American interests to reduce the monopoly power of individual primary product producers who perceive opportunities for profitable collusion after the oil embargo and the resulting price increases of fall 1973. An international organization of consuming nations would reduce that power. Yet, instead of consulting with the Japanese and perhaps proposing such a cooperative consuming-nation mechanism, unilateral actions in the soybean market greatly hurt our bargaining (and leadership) stance when the Arab oil embargo was announced, reducing the opportunities for collaboration among the consuming nations, and thus increasing the short-run degree of monopoly power gained by the exporting cartel.

III. Interdependence of American Economy and the Rest of the World

The effects of the international economy on the American are generally dealt with in the main chapters of the Report. Until recently, the international chapter was an appendage, and mention of the international economy in the main body of the Report was virtually nonexistent. Then, as American payments problems loomed larger, the international chapter seemed more important, but still largely unrelated to domestic issues. Finally, as Kenen noted two years ago, international policy objectives seemed to have overridden domestic in the August 15, 1971 measures.

This year's Report does not have the international tail wagging the domestic dog, but there does appear to be greater recognition of American interdependence with the rest of the world than has appeared in past analyses. Even in the President's transmittal message is the statement that among the "lessons" of 1973 is " . . . *the importance of the rest of the world.* The events of 1973 brought our external economic relations sharply to our attention. Most simply put, it will be ex-

[9] The futures market was constrained in the amount price could increase each day. Under that circumstance, future contracts exceeded (as in principle they always can and in fact often do) the available supply. To confound matters, American farmers were subject to price control on their beef, so predictably enough, soybeans were diverted to overseas buyers and political pressures to save the feeder lot operators from a price-cost squeeze were enormous. Some believe that removal of the limit on the future price alone would have solved the problem. It can equally plausibly be argued, however, that if price controls on beef were binding, as they were, then allowing free exports would have been inferior to *some* restriction on soybean exports in a second best world in the short run.

ceedingly hard for us to have a stable economy in an unstable world . . . " (p. 5).

Of course, it would be surprising if such recognition did not follow the boom in agricultural exports and the oil boycott. But recognition of interdependence permeates the entire Report, and goes far beyond the livestock-feed and oil situations. In terms of aggregate demand expansion in 1973, for example, " . . . a major element . . . " (p. 55) was the shift in the American net export position (a swing estimated at $11.5 billion from fourth quarter 1972 to fourth quarter 1973). Productivity growth was slowed partly by materials shortage, partly attributable to foreign demand and large exports (p. 63). Those exports, in turn, were larger than they would have been in the absence of price controls; controlled domestic prices with unrestrained foreign prices increased the attractiveness of the export market, thus intensifying shortages at home (p. 93). Inflation, too, is explained partly by foreign factors: import prices rose 26 percent, food price increases were accelerated by foreign factors (p. 67) and rising oil prices affected the price increases of the fourth quarter. Finally, the Energy and Agriculture chapter is of course devoted to discussion of issues closely related to the international economy.

In many ways, awareness of international economic interdependence may be the most important achievement of 1973, if it leads to more determined efforts to negotiate freer international economic relations. If, instead, the result is an attempt to reduce that interdependence to a considerable extent, the events of 1973 may mark the turning point toward restriction. To some extent, the appreciation of the American dollar in late 1973 and 1974 has intensified the risk, although it is too early to tell the extent to which the trade and current account balances will again shift in response to exchange rate changes.

IV. Coverage and the International Economic Report

In terms of the issues addressed in the Economic Report, the Council gives adequate coverage to the major issues of international economic policy, especially in view of the uncertainties that were present when the Report was written. That coverage is, by and large, more descriptive than analytical.

In an important sense, however, there is a serious omission: the economic reporting of what happened to prices and quantities of exports, imports, and other international transactions is woefully inadequate relative to the treatment given to domestic sectors of the economy. The reason for this probably lies in the historical concern of the Administration with issues pertaining to the American balance of payments. In the statistical section, there are only nine tables altogether, and the finest breakdown one gets of exports and imports is into the three-way split: food, beverages, and tobacco; crude materials and fuel; and manufactured goods. One table is on overseas loans and grants, i.e., our aid programs, one is a summary table on international reserves, and two give price indices.

Simultaneously with the release of the Economic Report, an International Economic Report of the President, written by the Council on International Economic Policy (*CIEP*), was issued. Much of it duplicates material covered in the Economic Report, although there is some additional material. There is, first of all, a set of fifty-four tables at the end, which have some interesting data. One such table provides the American trade balance in selected commodities from 1960 to 1973. Products with a rising trade surplus included nonelectric machinery, aircraft, computers, and basic chemicals. Products with a declining trade balance included motor vehicles and parts (from a $1.06 billion surplus in 1964 to a $3.8 billion

deficit in 1973), steel, textiles, and consumer electronics. The final eleven tables all give data on various aspects of the world petroleum market.

The trouble with the tables, as they stand, is that there are no sources given, and they are still nonoptimal in coverage. It would be a major undertaking to devise a satisfactory set of international tables (including, presumably, such statistics as prices, effective protection, domestic production, consumption, and trade in major commodity groups) to enable comparability between years and also permit the reader to go back to the data sources.

Other issues are also covered in the *CIEP* Report: an accounting of the major provisions of the Textile Arrangement (pp. 7–8); an explanation of the reconciliation of American and Canadian trade data (p. 34); a short chapter on Economic Aspects of the Law of the Sea; and New Approaches to Foreign Trade in Communist Countries. In a sense the *CIEP* Report appears to be geared more to business interests, as there are numerous colored graphs, and the level of economic sophistocation is very low:

> Economic theory dictates that investment, domestic or foreign, tends to have a positive impact on income, production, and employment in the country or region where the capital is put to work. . . . Foreign investment may substitute for, as well as supplement, domestic investment. To the extent that it merely takes the place of domestic investment, it does not increase the level of economic activity. Only that part of investment incremental to domestic investment has a noticeable effect on the domestic economy. . . . [p. 61]

Overall, the impression given by the *CIEP* Report is much more that of a corporation's annual report to its stockholders than that of a professional document.

There are a number of issues to be resolved, now that the balance of payments need no longer be the focus of the international chapter of the Economic Report; the existence of a separate *CIEP* Report adds further questions. There is a healthy opportunity for a shift in emphasis toward reporting on the welfare aspects of trade and investment policies. There is also a danger that focus on international aspects will be shifted from the Council's document to the *CIEP* Report. From the viewpoint of an academic economist, I would enter a plea for giving coverage to international economic activities in the main Report comparable to that given to other major activities, including an expanded and revamped set of international statistics. Inclusion of international aspects would enable preservation of the Economic Report as the useful overview of the economy by professional economists that it has been in the past. Splitting off reporting on aspects of foreign economic relations to the *CIEP* would downgrade international aspects, which would then fail to receive the attention of nonspecialists. It would also bely the apparent lesson of 1973; almost every phase of domestic economic activity is highly interdependent with the international economy; analysis of domestic policy issues without consideration of international aspects has become a virtual impossibility.

REFERENCES

R. E. Baldwin, *Nontariff Distortions of International Trade*, Washington 1970.

R. N. Cooper, *The Economics of Interdependence: Economic Policy in the Atlantic Community*, New York 1968.

P. B. Kenen, "The 1972 Report of the President's Council of Economic Advisers: International Aspects," *Amer. Econ. Rev.*, Sept. 1972, *62*, 517–32.

M. Parkin, "The 1973 Report of the President's Council of Economic Advisers: A Critique," *Amer. Econ. Rev.*, Sept. 1973, *63*,

The 1974 Report of the President's Council of Economic Advisers: The Control of Inflation and the Future of the International Monetary System

By DAVID LAIDLER*

In this essay I shall comment on two issues raised in the Economic Report of the President: the control of inflation and the reform of the international monetary system. In the domestic sphere I shall argue that wage and price controls failed because the conduct of monetary policy was inconsistent with their success. In the international sphere I shall argue that the real issues involved in reforming the international monetary system concern the extent to which individual countries can be expected to coordinate their domestic stabilization policies—particularly on the monetary side.

To the author the most striking section in the Economic Report is the commentary on the breakdown of wage and price controls. Though it is argued that these controls had some impact on the inflation rate in 1971 and 1972, the conclusion as far as 1973 is concerned is that " . . . the situation the program ran into was more than it could successfully contend with" (p. 103). Though unwilling to " . . . rule out the possibility that inflation might have been even greater in 1973 without controls" and pointing out that " . . . no one can disprove the thesis that controls had a significant effect," the Council nevertheless concede that " . . . 1973 makes it a hard thesis to believe" (p. 108). To a British commentator the willingness to admit failure so openly is commendable indeed. The contrast with official British attempts to defend a similar system of controls whose breakdown was much more dramatic and obvious is, to say the least, vivid.

Nevertheless, commendably honest though the Report is on this question, it is much less helpful in explaining why the controls broke down. What exactly was "the situation that the program ran into"? The ingredients to an answer to this question are in the Report, but they are not put together coherently. This failure stems not so much from lack of honesty as from faulty analysis. To the author it seems that the Council both seriously underplay the importance of the behavior of the money supply in generating inflation, and, closely connected, considerably underestimate the length and complexity of the time lags implicit in the transmission of monetary effects.[1] As a result they fail to see that the "situation" that wage and price controls "ran into" was no more than the predictable consequence of an expansionary monetary policy.

In 1969 the money supply (M_2) grew by 3.4 percent, in 1970 by 7.5 percent, in 1971 by 11.4 percent, in 1972 by 10.8 percent, and in 1973 by 8.8 percent.[2] The Report in

* University of Manchester. I am indebted to Michael Parkin and George Zis for many helpful comments on an earlier draft of this essay and to the SSRC for financial support.

[1] This year's Report is not unique in this respect. See Reuben Kessel and David Meiselman for similar comments on earlier Reports.

[2] These data are based on Table 18, p. 82 of the Re-

more than one place refers to a two-quarter lag between the rate of change of M_2 and the rate of change of nominal income, and it is not easy to reconcile so short a lag as this with an inflation rate that accelerated rapidly in 1973 and promises to continue to accelerate at least until mid-1974. However, if one instead takes two years as being a rough estimate of the time that it takes for the rate of monetary expansion to work its way through to the inflation rate, then there is no problem in understanding the data.

Two questions suggest themselves here: first, is it appropriate to treat the rate of monetary expansion rather than the level of the money supply as the key monetary policy variable? Second, is it reasonable to postulate so long a time lag? There is good reason, both theoretical and empirical, for taking an affirmative answer to both questions seriously. The following simple model, which I have discussed in detail elsewhere (1973), essentially combines the quantity theory of money with a version of the expectations augmented Phillips curve of Milton Friedman and Edmund Phelps; one of its key implications is that the rate of monetary expansion is the monetary variable upon which one should focus in discussing policy.

Let M denote the natural logarithm of money, Y^* the natural logarithm of capacity real income (assumed to grow at a constant rate), y the natural logarithm of the ratio of actual to potential output, and hence an excess demand proxy. Y is y plus Y^* and hence the logarithm of real output, and P is the natural logarithm of the general price level. The superscript e denotes the expected value of a variable, a bar over a variable denotes its exogeneity and subscripts -1 and -2 denote time lags; and Δ denotes a first time difference (so that the model effectively is written in terms of proportional changes in variables). We have:

$$\Delta\overline{M} = b\Delta Y + \Delta P \tag{1}$$

$$= b\Delta Y^* + b\Delta y + \Delta P$$

$$\Delta P = gy_{-1} + \Delta P^e_{-1} \tag{2}$$

$$\Delta P^e = d\Delta P + (1-d)\Delta P^e_{-1} \tag{3}$$

Equation (1) is simply a version of the quantity theory of money, while equation (3) is the familiar "adaptive expectations" formula applied to the inflation rate.[3] Equation (2) is a price formation equation that implies that firms set the rate of increase of the price of their output equal to the expected rate of inflation but modify their pricing decisions in the face of excess demand or supply for goods.

An expression for y, the excess demand proxy in this model, is given by

$$y = \frac{1}{b}\Delta^2 M - \Delta^2 Y^* + \frac{2b-g}{b}y_{-1} - \frac{b-(1-d)g}{b}y_{-2} \tag{4}$$

As the reader will see, it is *changes* in the percentage rate of monetary expansion that affect the level of excess demand, but this result does *not* depend upon the model being set up in first difference form, nor does it rely on the absence of an opportunity cost of holding money variable from equation (1). It follows instead from the role played by inflationary expectations in the price formation equation (2). When inflation is going on, a certain rate of

port itself. Note that the rate of expansion of M_1 actually continued to accelerate slightly in 1972.

[3] If the expected rate of inflation were included in the demand for money function as a measure of the opportunity cost of holding money, the model's basic behavior would not be altered. Moreover, if we assume that the banking system pays interest on its deposit liabilities at a competitive rate, equation (1) may in any event be defended as a reasonable formulation of the demand for money function, particularly if its empirical implementation involves the use of M_2 as the definition of money, which in fact it does. See the author (1973).

nominal monetary expansion is required to keep the supply and demand for money in equilibrium and aggregate demand at a level compatible with capacity real output. Hence changes in the rate of change of the nominal money supply cause monetary disequilibrium rather than changes in its level. If this model is regarded as at all reasonable then it is also reasonable to regard the rate of nominal monetary expansion as the key monetary policy variable.[4]

As to the lag in the effects of monetary policy, there is already a substantial literature on the matter, and there is no need to survey it here.[5] Suffice it to note that, no matter what technique has been used to investigate it, the lag between monetary changes and aggregate demand has always been found to be "long"—a period of between six months and a year usually elapsing before a significant impact is to be observed. Moreover, in the present context we are concerned not with the impact of monetary changes on aggregate demand but on the inflation rate. A key implication of the expectations augmented Phillips curve used above is that changes in aggregate demand have their initial impact on real output and employment and only subsequently affect the inflation rate.

According to the relevant analysis, output and employment changes provide the signals to firms that indicate the desirability of altering the rate at which they change both the prices of their output and the money wages that they offer to both actual and prospective employees. Unless output and employment differ from their capacity level, the rate of inflation will be equal to that anticipated rate (see equation (2)) and will not change unless expectations change exogenously. It is not *a priori* implausible that this stage in the transmission process can take as long as a year, and there is some support for this view implicit in empirical work based on the model set out above. Equations (2) and (3) yield

$$(5)\quad \Delta P = gy_{-1} - (1-d)gy_{-2} + \Delta P_{-1}$$

and when fitted to annual data for the United States for the period 1953–72, this equation performs remarkably well.[6]

If the time lags inherent in the *U.S.* economy are as long as I have suggested and if the rate of monetary expansion is indeed the key variable determining the behavior of aggregate demand, then the breakdown of wage and price controls was quite predictable. The Report puts the matter this way: "The main object . . . was to slow the rate of inflation. It would be achieved, in part, by postponing some price increases in the expectation that monetary and fiscal policy would slow the rate of growth of demand and perhaps make them unnecessary" (p. 97). But controls were in force from August 1971 onwards and 1972–73 was the very period in which the effects of monetary policy on the inflation rate should have been expected to be significantly expansionary. Only in 1974 should we expect to see the first effects on the inflation rate of the slow monetary contraction that perhaps began in 1972 and continued into 1973.

It is worth looking at what happened over the last few years in a little more detail, if only to bring out clearly how difficult a task it must necessarily be to coordinate wage and price controls with mone-

[4] Thus the somewhat inconclusive debate between John Culbertson (1960, 1961) and Friedman (1961) on this question would seem to be settled in Friedman's favor by the foregoing analysis.

[5] Thomas Mayer, especially ch. 6, provides a useful summary of much of the earlier work on this issue. Note that the lag under discussion here is concerned with impact effects only and not with longer run questions about the time path followed by the economy after the initial impact. The model just discussed abstracts entirely from this lag for the sake of simplicity.

[6] In particular a unit coefficient on the lagged inflation rate was found. See the author (1973) for a full discussion of these results.

tary policy. The object of using such controls is to bring down the rate of inflation while reducing the costs in terms of unemployment and lost output that would be incurred by relying solely on demand management policies.

Implicit in the arguments already presented above is the proposition that once they are built into the economy, inflationary expectations can only be removed by having excess supply in the system which results in the actual inflation rate falling below that anticipated; the current behavior of the inflation rate determines the future course of the expected inflation rate. In the context of this view, the introduction of wage and price controls perhaps acts exogenously on inflationary expectations, reduces them, and hence reduces the actual inflation rate without the intervention of excess supply in the economy. At the same time, however, aggregate demand must be reduced in order to ensure that its time path is compatible with the new lower inflation rate. This step is vital, for just as deficient demand slows the inflation rate down below its expected value, so excess demand raises it above that value. There is no point in exogenously reducing inflationary expectations if the behavior of aggregate demand leads to their being disappointed and hence revised upwards again. Only if a decrease in inflationary expectations induced by the introduction of wage price controls can be coordinated with appropriate demand management can one cut the inflation rate without the intervention of a fall in income and employment. This does not mean that the inflation rate is cut down costlessly, for wage price controls inevitably result in losses due to allocative inefficiencies, but there does seem to be a widespread view that these are more tolerable than those implicit in the deliberate generation of unemployment.[7]

TABLE 1—U.S. GOVERNMENT SECURITY YIELDS 1971–73

	3-Month Treasury Bills	3–5 Year Issues
1971		
June	4.7	6.4
August	5.1	6.4
October	4.5	5.7
December	4.0	5.4
1972		
February	3.2	5.5
April	3.7	6.0
June	3.8	5.8
August	4.0	5.9
October	4.7	6.1
December	5.1	6.1
1973		
February	5.6	6.6
April	6.3	6.7
June	7.2	6.8
August	8.7	7.8
October	7.1	6.8
December	7.4	6.8

Source: Table C.58, p. 318 of the Report.

Certain evidence points to the Nixon Administration's wage and price controls having successfully reduced at least short-term inflationary expectations and hence, for a while, having reduced the actual inflation rate. However, the policy failed utterly on the crucial question of being coordinated with demand management. It is well known that the end of 1971 saw the beginning of a sharp fall in nominal rates of interest, a fall that was particularly marked at the short end of the term structure. From the beginning of 1972 onwards this fall reversed itself and by the end of that year rates were more or less back to their old levels. By the end of 1973 they were higher still. Table 1 presents some relevant evidence. It is plausible in the extreme that the behavior of interest rates

[7] The foregoing argument emphasizes the role played by expectations in the formation of prices on what we might term the "supply" side of the economy. Though recognizing that expectations have a role to play in the inflationary process, the Report stresses their effects on demand. Perhaps the influence of the expected rate of inflation on the velocity of circulation forms the centerpiece of the Council of Economic Advisers' analysis, but the Report could be clearer than it in fact is on how controls are supposed to affect the inflation rate. See in particular, pp. 100–01.

summarized there partly reflects an expectation on the part of the public that the wage price control program would significantly reduce the inflation rate in the short term. Interest rates, particularly short rates, fell by more than would have been justified by the actual fall in the inflation rate before August 1971. However, the expectation that wage-price controls would succeed was systematically eroded during 1972 and 1973.[8]

As to the effect of wage and price controls on the actual inflation rate, we do not have such clearcut evidence. For example, Edgar Feige and D. K. Pearce argue that the controls had no effect on the course of inflation. However, there is work (for example, Robert Gordon) that suggests that the policies did at first lower the inflation rate, and empirical implementation of a version of the simple model set out above yields similar implications. The model was modified both to allow for a lag in the effect of monetary expansion on real income and to permit the inflation rate to respond with a distributed lag, rather than a one-period delay, to excess demand.[9] In this form it was fitted to annual data for the period 1953–65 and, with all the a priori constraints on parameter values implicit in its structure imposed, it yielded the following results (with t-values in parentheses).

$$(6)\quad \underset{(1.495)}{\Delta Y=.493(\Delta M-\Delta P)}+\underset{(2.037)}{.735(\Delta M-\Delta P)_{-1}}$$

$$(7)\quad \Delta^2P=\underset{(3.018)}{.207y_{-1}}-\underset{(2.543)}{.168y_{-2}}+\underset{(1.871)}{.208\Delta^2P_{-1}}$$

When these parameter values were used in a simulation exercise which employed starting values of the lagged endogenous variables for 1951–52 but thereafter permitted the model itself to generate them, it produced a set of predictions for the rate of change of the inflation rate as measured by the GNP deflator for the period 1966–73 which are reproduced along with actual values in Table 2.[10] The size of the predictive errors for 1972 and 1973 are larger than those for earlier years and their signs suggest that although wage and price controls perhaps reduced the inflation rate in 1971–72, they produced a strong "catch-up" effect in 1972–73. Thus, the total effect of controls, according to these results, was temporarily to slow down and then temporarily to accelerate the inflation rate rather than permanently to reduce it. If the rate of monetary expansion is significantly increasing the rate of growth of aggregate demand, as it certainly was in 1972 allowing for up to a one-year lag, at the same time as wage and price controls are slowing down the actual and expected rate of inflation, then excess demand for goods must be building up which will eventually make its effects felt on prices. These effects were felt in 1973 and are continuing to be felt at the beginning of 1974.

TABLE 2—ACTUAL VS. PREDICTED CHANGE IN THE INFLATION RATE 1966–73[a]

	Actual	Predicted
1966–67	0.4	1.1
1967–68	0.8	0.9
1968–69	0.8	0.7
1969–70	0.6	0.5
1970–71	−0.7	−0.3
1971–72	−1.6	−0.3
1972–73	2.1	0.5

[a] Second difference of natural *log* of *GNP* deflator. Original variable measured at midyear. Thus the figure for 1972–73 measures the difference between the inflation rate prevailing from June 1972 to June 1973 and that prevailing over the preceding year.

To make wage and price controls work, according to the above analysis, it would have been necessary to time the effects of monetary policies so that they coincided

[8] Note that Michael Parkin drew attention to this evidence in discussing last year's Report.

[9] See the author (1973) for a full account of this.

[10] Note that the model tended systematically to overpredict the inflation rate by a little over one percentage point over these years.

with the reduction in the inflation rate brought about by wage and price controls. The effects of the two policy tools could then have been mutually reinforcing rather than contradictory. However, given the time lags involved, the relevant monetary policy should have been undertaken a year or more in advance of the introduction of controls to achieve even a rough and ready degree of coordination. As it was, the thrust of the effects of monetary policy during 1972 and 1973 was in the opposite direction to that of wage and price controls, and it is hardly to be wondered at that the anti-inflation policy failed.

As things now stand, one cannot argue with the Council's view that " . . . while continued rapid inflation is not inevitable, the course of unwinding it will be long and difficult" (p. 21). Inflationary expectations are now deeply embedded in the *U.S.* economy and will not be reduced again by a new round of controls. The impact of controls on inflationary expectations depends crucially upon the acceptance on the part of the public that they will succeed; once they have been seen to fail, that acceptance vanishes.[11] Demand management policies are all that seem to be available and past experience suggests that the economy's response to them might be slow indeed. The results of a simulation exercise carried out with the model set out in equations (6) and (7) suggest that, were the rate of monetary expansion to be cut by as much as 4 percentage points between 1973 and 1974 and thereafter held constant, the inflation rate would not fall below 3 percent till 1979, even though real income would actually fall in each of the three years 1975–77 with its rate of growth not getting back to its assumed long-run rate of 3.6 percent till 1980.[12]

These results are, however, probably on the pessimistic side, even if the model from which they are derived is taken absolutely seriously as a forecasting device; and the reader will undoubtedly be somewhat sceptical about so simple a model being able to provide any serious conditional forecast of the behavior of the *U.S.* economy. Nevertheless, however the reader interprets these results, the slowness of the model's response to monetary contraction and the large quantity changes required to induce changes in the inflation rate have a common source: namely the slowness of inflationary expectations to respond to experience implicit in the parameter estimates upon which the simulation was based. These estimates were derived from the relatively noninflationary period 1953–65. There is every reason to believe that the American public have become more inflation conscious over the last few years and hence quicker to adapt their expectations to experience, as the Report itself notes.

Perhaps more important, but not unrelated to a growing awareness of the inflationary problem, has been an increasing tendency for wage settlements to include cost-of-living escalator clauses. This tendency is unremarked in the Report, but has surely had something to do with the lack of industrial strife that has accompanied accelerating inflation in the United States, a feature of the American scene which sharply differentiates it from events in Europe where even West Germany is

[11] It is worth noting though that the announcement of any anti-inflation policy in which the public believes will lower inflationary expectations. There was never anything unique about wage and price controls in this respect. I discussed this matter in detail in my 1971 article.

[12] Moreover this exercise does not have the economy converging closely on a steady-state solution until well into the 1990's and that convergence is cyclical. The period implicit in the model which must elapse before anything approaching full adjustment to a new rate of monetary expansion is over two decades—just the period, incidentally, which Friedman (1968) suggested was required for the *U.S.* economy fully to adjust to such a change.

suffering a little from what used to be known as "The English Disease." Quite apart from its probably beneficial effects on the industrial relations front, the adoption of escalator clauses increases the rate at which demand management policies can change the inflation rate (either upwards or downwards). Hence, it cuts the potential cost in terms of unemployment of slowing down the inflation rate at any particular pace. Without escalator clauses, deficient demand leads to unemployment; this decreases the inflation rate which in turn decreases inflationary expectations so that future wage and price inflation rates are also cut. Each step in the process takes time, but the use of escalator clauses "short circuits" the expectations adaption stage and hence speeds up the whole process.[13] None of this, of course, means that the control of inflation in the United States is going to be easy, or that it is going to be costless; only that it is going to be a little easier than it might otherwise have been, and that it will cost a little less.

I have argued that the Economic Report is less coherent than it might be in dealing with domestic policy towards inflation. It can similarly be faulted for its treatment of the international monetary system. Again, this vagueness seems to this writer to stem from a relative neglect of monetary influence on income, employment, and prices. Chart 2 on page 56 of the Report shows changes in real *GNP* of the United States, Japan, and OECD Europe. From 1961 to 1968, only one pair of years (1963–64) saw growth rates vary in the same direction for all three. From 1969 onwards, growth rates have always moved together year by year. This synchronization of the business cycle across industrial countries has obviously had a great deal to do with the behavior of commodity prices over the last year or so, for quite apart from inflation induced speculation, we have just been through a cyclical upswing in which everyone was building up his inventories. However, more interesting for present purposes than the effects of this phenomenon are its probable causes.

If we rule out simple coincidence as an explanation, we must look for a factor that has led to a marked increase in the interdependence of national economies during the 1960's. The prime candidate here is the recent rapid growth of a large scale market in internationally mobile capital. The effect of this development has been to render it impossible for any country linked to the United States by a fixed exchange rate to have a domestic monetary policy independent of that being conducted in the United States. The year 1969, when cycles became synchronized, was, after all, the year in which a tight monetary policy in the United States combined with the effects of Regulation Q to produce sharp increases in interest rates and a slowing down of monetary expansion rates in much of the industrialized world. The subsequent rapid reversal of *U.S.* policy coincided with widespread adoption of controls on capital movements, and the eventual breakdown of the Bretton Woods system of fixed exchange rates as individual countries tried to reestablish domestic control over their monetary policies.[14]

The Report only goes part way to recognizing the problems involved here. It does pay a little attention to the pressures that capital movements put on countries whose exchange rates are out of equilibrium (p. 206), and does suggest that the future of the international monetary system must make allowance for greater exchange rate flexibility (how much more it does not

[13] There has been recent widespread discussion of indexing and escalator clauses. My treatment here of the problems involved owes much to conversations with Herbert Giersch, David Grove, Richard Jackman, and Michael Parkin.

[14] A lucid account of these developments is to be found in Bell.

say), (p. 204), but it devotes considerable space to discussing the role of primary reserve assets in the future of the international monetary system. As the Report itself says: "The issue of convertibility into primary reserve assets arises primarily in the context of an agreement to limit fluctuations of exchange rates by government intervention in the foreign exchange market" (p. 206) and it notes that "The Committee of Twenty has agreed that in the context of long-term monetary reform it would be desirable to establish the convertibility of . . . currencies into primary reserve assets insofar as it is agreed to stabilize currencies within agreed margins" (p. 207).

Some form of fixed exchange rate system, then, is still on the menu as far as the future of the international monetary system is concerned. The Report does not face up to the implication of choosing such an alternative for the conduct of domestic policy. The original Bretton Woods system was premised on the fact that when exchange rates were temporarily out of equilibrium, government intervention in foreign exchange markets could maintain a disequilibrium rate long enough for domestic policy actions to be taken in order to restore external balance at the existing exchange rate. The system attempted to have the best of two worlds: it tried to combine the undoubted advantage accruing from something akin to a unified world currency with the advantage of a system that permitted individual countries to manage in the short run their own levels of prices, income, and employment.

Under this system, so long as the balance of payments was dominated by trade and long-term capital movements, the obligation to maintain a fixed exchange rate imposed a loose, long-run coordination on the economic policies of individual countries. As short-term capital movements became more and more important, the time period over which official intervention in the foreign exchange market could sustain a disequilibrium exchange rate became shorter and shorter, the change here being particularly marked for surplus countries. Thus the degree of coordination of domestic policies required to maintain fixed exchange rates became closer and closer. Such coordination was effectively imposed on domestic monetary policies by the very fact that it became impossible for any country to maintain a level of interest rates out of line with those required for international capital market equilibrium.

Now the implication of this for any future official attempts to prevent exchange rates fluctuating outside predetermined boundaries is quite straightforward. The trick can be worked only if domestic policies are so coordinated with those being carried out elsewhere as never to produce the market forces that would take the exchange rate outside those bounds in the first place. The coordination of even short-run domestic policies is probably now prerequisite for maintaining stable exchange rates. If policies are so coordinated, exchange rates will be stable without official intervention; if policies are not so coordinated, then the large international capital market will ensure that official intervention will not be capable of stabilizing exchange rates.

Thus the effective choice for the future of the international monetary system is not simply one between fixed and flexible exchange rates. Rather it is a choice between having a high degree of coordination between national economic policies under stable exchange rates, and leaving individual nations free to pursue independent domestic policies where exchange rates vary to the extent that such policies diverge. The Economic Report is vague

about the future of the international monetary system precisely because it never faces up to the fact that the choices involved in its reform concern not just international cooperation about exchange rate policies, but the far more difficult question of international coordination of policies towards domestic income, employment, and prices. Given the key role played by interest rates on the international scene, stable exchange rates require close coordination of monetary policies.

Having said all this though, the Report's very vagueness on the matter of the future of the international monetary system gives reason for a certain amount of optimism. After all, the United States made an important contribution to international monetary stability by eliminating its deficit in 1973, and there is no sign in the Report that the current Administration is likely to press for any rapid attempt to restore fixed exchange rates. The fact that the Council of Economic Advisers can conclude that during 1973 "In general . . . foreign exchange markets functioned remarkably well" suggests that the system might be left to itself in the foreseeable future. Those who have long advocated exchange rate flexibility should welcome such neglect, for so long as no one is pressing for quick reform of the international monetary system, we will have exchange rate flexibility, albeit imperfect, as the only available means of organizing the international monetary system. Hence we will at least have an international monetary system capable of withstanding the strains imposed by the efforts of individual countries to bring their inflations under control at different rates and perhaps by different means.

REFERENCES

G. Bell, *The Euro-Dollar Market and the International Financial System*, London 1973.

J. M. Culbertson, "Friedman on the Lag in the Effect of Monetary Policy," *J. Polit. Econ.*, Dec. 1960, *68*, 617–21.

———, "The Lag in Effect of Monetary Policy: A Reply," *J. Polit. Econ.*, Oct. 1961, *69*, 467–77.

E. L. Feige and D. K. Pearce, "The Wage Price Control Experiment: Did it Work?," *Challenge*, July-Aug. 1973, *16*, 40–44.

M. Friedman, "The Lag in Effect of Monetary Policy," *J. Polit. Econ.*, Oct. 1961, *69*, 447–66.

———, "The Role of Monetary Policy," *Amer. Econ. Rev.*, Mar. 1968, *58*, 1–17.

R. J. Gordon, "Wage Price Controls and the Shifting Phillips Curve," *Brookings Papers*, *2*, Washington 1972, 385–421.

R. A. Kessel, "The 1972 Report of the President's Council of Economic Advisers: Inflation and Controls," *Amer. Econ. Rev.*, Sept. 1972, *62*, 527–32.

D. Laidler, "The Phillips Curve, Expectations, and Incomes Policy," in H. G. Johnson and A. R. Nobay, eds., *The Current Inflation*, London 1971.

———, "The Influence of Money on Real Income and Inflation: A Simple Model with some Empirical Tests for the United States 1953–72," *Manchester Sch. Econ. Soc. Stud.*, Dec. 1973, *41*, 367–95.

T. Mayer, *Monetary Policy in the United States*, New York 1968.

D. I. Meiselman, "The 1973 Report of the President's Council of Economic Advisers: 'Whistling in the Dark'," *Amer. Econ. Rev.*, Sept. 1973, *63*, 527–34.

M. Parkin, "The 1973 Report of the President's Council of Economic Advisers: A Critique," *Amer. Econ. Rev.*, Sept. 1973, *63*, 535–45.

E. S. Phelps et al., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

The 1974 Report of the President's Council of Economic Advisors: A Critique of Past and Prospective Economic Policies

By FRANCO MODIGLIANI*

I. How Good a Year Was 1973?

Judged by traditional standards, the year 1973 might appear as a success story for economic policy makers. Real *GNP* rose by just about 6 percent relative to 1972, a rate of growth right at the midpoint of the rosy forecasts made at the beginning of the year which clustered unusually narrowly around 6 percent; employment rose by well over 3 percent; productivity (in the private sector) by nearly 3 percent; and hourly wages (adjusted for overtime and industry shifts) rose a relatively modest 6.2 percent, the lowest figure since 1968 and almost puny when compared with the experience of many other industrial countries. Even the *GNP* deflator increased but 5.5 percent, clearly outside the fashionable forecasted range of 2.5 to 4 percent, but still not outlandish in this era of worldwide inflation, quite commonly at the two-digit rate.

Needless to say, these figures, the kind one would probably hear widely quoted if this were election time, are extremely misleading; they largely reflect the total inadequacy of the customary year-over-year comparisons. The truth is that, by most standards, 1973 was more nearly a debacle than a success story for economic policy, with the country gripped by the worst rate of inflation of any postwar peace year, with many commodities hard to get at times, both for consumers and for producers, with temper-raising gas queues a frequent sight at the turn of the year, and the dollar—no longer as sound as a dollar!

Measured from fourth quarter to fourth quarter (the standard that I shall use hereafter unless otherwise stated),[1] the rise in the *GNP* deflator was 7.1 (7.7 if measured by a fixed weight price index, Table C-4); but wholesale prices rose 17 percent—more than 30 percent for farm products and processed food and 12 percent for industrial commodities—and the cost of living by 8.4 percent. Furthermore since the increase in hourly wages during 1973 itself was only 6.7 percent, real hourly wages adjusted for overtime and industry shifts *declined* rather steadily throughout the year by a total of nearly 2 percent, and

* Massachusetts Institute of Technology. I am indebted to my colleagues, Stanley Fischer, Robert Merton, Paul Samuelson, and Lester Thurow, for reading a preliminary draft and making valuable suggestions, as well as for the opportunity I have had to discuss with them, on numerous occasions, many of the issues analyzed in this review. I also wish to express my thanks to Albert Ando for help in designing the simulations reported in Section IV and in interpreting them.

[1] This measure, it should be acknowledged, is not without shortcomings of its own, mainly that it is independent of the path followed by the variable within the year. One might remedy this shortcoming by taking as a measure the average quarterly growth above the last quarter of the previous year (which is equivalent to a weighted average of the quarterly growth, weighted 4, 3, 2, 1). For my present purpose, however, I feel that the last quarter to last quarter measure is the most suitable, though recognizing that, because the growth of *GNP* was by far largest in the first quarter, this measure may somewhat underrate the performance of the year.

real weekly earnings declined even more, by 3.1 percent. This is a performance without precedent on similar scale in the postwar *U.S.* economy, except possibly for 1951. Nor, surprisingly enough, were these soaring prices and reduced real wages the hallmark of a "capitalist paradise." There was no great leap in corporate profits, which in real terms increased only 5 percent, while total property income (profits plus interest plus rents) increased a much more modest $3\frac{1}{2}$ percent—an apparent puzzle which will be resolved shortly. And, to boot, despite the rise in profits, the stock market, reflecting among other things the deep anxieties of the country, lost 20 percent in the course of the year. Since there was but one important exogenous shock in 1973 that could not reasonably have been anticipated, namely the so-called Arab oil embargo, and that came only at the tail end, one would be led to conclude that the disastrous performance outlined above was largely the result of an ill-conceived and ill-implemented economic policy.

In advance of reading the Economic Report of the President I had been inclined to this conclusion, attributing the poor record to four major policy mistakes: i) an overexpansionary aggregate demand policy, both fiscal and monetary in 1972, especially in the last half of the year, and in early 1973; ii) inadequate provisions for expanding agricultural output after the sales to Russia and China; iii) encouraging or at least tolerating the depreciation of the dollar, especially after the February devaluation and through the third quarter; iv) the retention, and even more, the poor management of an extensive system of controls on prices, wages, etc.

Holding these views, which I had partly expressed in print,[2] I looked forward eagerly to study the Council's Report to see whether it would provide grounds for modifying these negative judgements, or explanations of what had caused the errors.

II. The Council Account of Developments and Policy in 1973

A study of the Council's account of developments and aggregate demand management in 1972–73 seems to me to provide little ground for modifying my prior view that the Administration and the Council managed to repeat the terribly costly error committed by the Democratic Administration in the mid-1960's, namely that of permitting the rate of expansion of the economy to accelerate as we were getting closer to full utilization of resources. In 1963, when unemployment stood at an average of 5.7, the rate of growth of *GNP* was but 4.4; in 1964 with unemployment down to 5.2, it was 4.5, but in 1965, which opened with unemployment down to 5 percent, growth accelerated to 8.2 percent bringing unemployment to the critical area of 4 percent; and, as that line was being crossed at the turn of the year, the growth was allowed to leap to over 9.4 percent in the fourth quarter and 8 percent in the first quarter of 1966. Lack of resources finally slowed down the growth in the rest of the year to below 4 percent, but at the cost of putting pressure on prices which in some months grew at a rate approaching 10 percent at wholesale and 6 percent for the cost of living. This opened up the inflationary era from which we are still unable to extricate ourselves.

Ever since that disastrous experience I had been in dread of a "second," and as the Nixon Administration game plan developed, my worries kept increasing. By July of 1971, the Administration had brought unemployment back to the 6 percent area, income was growing at a sluggish 3 percent or so, yet no fiscal actions were being taken to speed the recovery and the monetarists were giving hell to the Fed-

[2] Notably in a series of articles in the Milan, Italy, newspaper *Corriere Della Sera*.

eral Reserve because the money supply in the previous six months had grown at over 12 percent. At that time, testifying before the Joint Committee on the Economic Report, I pleaded for more stimulating policies, explaining that, given the fiscal inaction, there was nothing wrong with a 12 percent growth of M_1 at that time, and that, anyway, the M_1 series, and especially the seasonal adjustments, were too unreliable to warrant paying attention to every wiggle of the annualized rate of growth of M_1,[3] and finally expressed my alarm in the following rather frank terms: "What I am concerned with is the following thing, that if we have a slow recovery in 1971, as we seem to have, and we are approaching election time, there will be all kinds of measures taken to stimulate the economy quickly at that time, to be sure we get through the election time with a reasonable level of employment, and then we will be picking exactly the wrong tactic; namely, going slowly when we are far away from the full employment goal, and running quickly when we get close to the goal, with a large chance to overshooting it" (1971, p. 123).

I submit that this statement came pretty close to anticipating correctly one of the major causes of trouble in 1973 (though I shall leave it to the reader to decide whether my suggested motivation was the relevant one!). In 1972, as unemployment was moving rapidly down, the rate of growth accelerated to 7 percent; in the last (and election) quarter, with unemployment down to just over 5 percent, the rate of growth exceeded 8 percent.

All this was achieved through a monetary-fiscal policy mix which was extraordinarily expansive throughout the year and became even more so at the year's end. The actual federal government budget deficit on *NIA* account was $16 billion, a figure which further rises possibly by as much as $9 billion, if one accepts the suggestion of the Economic Report (pp. 29–30) that, in terms of effect on consumers' expenditure, overwithholding in 1972, estimated at some $9 billion, should be primarily treated as an addition to the stock of claims of the private sector against the government, rather than as ordinary tax. With this correction, the Council estimates the full employment deficit in 1972 at $7 billion, using the conventional measure of "full employment" (4 percent unemployment) and at over $10 billion using a variable measure of "full-employment unemployment" corrected for labor force composition, as now recommended fairly persuasively by the Council (p. 30 and Table 1). If one does not accept fully the overwithholding correction, then the full-employment deficit would be lower, possibly, in the limit, down to $1.5 billion; however, since the Council seems to regard close to full correction as appropriate, it could hardly fail to regard 1972 fiscal policy as highly stimulative. Furthermore, in the year's last quarter, a sudden spurt of expenditure brought the actual deficit to an annual rate of $23 billion, implying a full-employment deficit of $19 billion (presumably corrected for overwithholding, Table 17, p. 80). To be sure, some $10 billion of this increase represented the first installment of the new revenue sharing system, and was therefore partly offset by an increase in the surplus of state and local governments; still the combined full-employment surplus for all governments, as estimated by the Council, fell by some $11.5 billion (Table 17, last column).

As for monetary policy, throughout 1972 money (by which I always mean M_1), grew at a rate in excess of 9 percent. But, what is to me more significant is that long-term interest rates, which had been al-

[3] Since that time the estimate of M_1 for that period has been increased by some $6 billion, or 3 percent, and the implied rate of growth for the first six months has been revised down from 12.3 to 10.9 percent.

lowed to rise in the first half of 1971 when unemployment was holding steady at 6 percent, were still falling, or at their lowest point, at the turn of 1972. Thus the mortgage rate, as measured by the FHA new home mortgage yield, stood at 7.57 as compared with nearly 8 in mid-1971 and the AAA new issue rate, computed by the Federal Reserve, stood at around 7.1 as compared with about 8 percent. It seems hardly surprising then that, in the first quarter of 1973, which opened with unemployment down to 5 percent, the growth rate of real *GNP* accelerated further to a mammoth 8.6.

My hunch that such an outlandish rate of expansion—roughly twice the long-term rate of growth—was highly inappropriate under the circumstances, much as the corresponding rate in the first quarter of 1966, receives strong support from some simple calculations exhibited in Table 7 of the Report. These calculations are based on the consideration, by now widely accepted by students of the unemployment syndrome, that for a given state of demand pressure—as measured, say, by overall vacancies—demographic groups of different age, sex, race, etc., may have different specific rates of unemployment (because of differential frequency of quits or layoffs, or entering and leaving the labor market, for example). If, then, the composition of the population changes significantly over time, the same state of demand will imply a different overall rate of unemployment; in particular, if the composition shifts toward groups with higher specific unemployment rates, as has happened in the United States since the mid-1950's, then the overall rate will rise, and the target full-employment level of unemployment must be raised accordingly.

One simple way to make a rough allowance for changes in composition is to compute a "standardized" unemployment rate for every year, by weighting the unemployment rate of each group in that year by a standard labor force composition, say that of a specific year, which thus serves as a sort of base year. Table 7 shows that if one uses as the base year 1956, in which unemployment was 4.1 percent, or very close to what was considered then the full-employment rate, then in 1972, the actual unemployment rate of 5.6 percent is equivalent to a standardized rate of only 4.9, and the actual 1973 rate of 4.9 percent is equivalent to a standardized rate of but 4.1 percent—the same as in 1956. This is clearly a striking result. It confirms my view that the end of 1972, when the actual unemployment rate was down to 5.3 percent and the corresponding standardized rate down to but 4.6 percent, was a time calling for a very moderate rate of expansion of aggregate demand, not significantly higher than the long-run growth of 4 to 5 percent.

It should be recognized at this point that the Council's calculations might tend to overstate the tightness of the labor market in 1972–73 relative to 1956; the reason for this conclusion is that, in the two years the group-specific rates of unemployment were by no means identical. Specifically, in both years the rate of unemployment (hereafter U) for the prime groups, male 20 years and over, was the same or lower than in 1956; but the U for all other age groups (except females over 55) was higher, sometimes appreciably. Under these conditions, the results are not invariant under the choice of the base year. In particular, since in 1956 the prime group was relatively more abundant than in the later years, if we use 1973 as the base year, the standardized U for the more recent year will rise relatively to that of 1956. Thus, using 1973 weights the standardized U for 1956 is found to be 4.6, instead of the actual 4.1 (see fn. 3 to Table 1), which is somewhat lower than the actual rate in 1973, namely 4.9. In other words, in terms of 1973 weights the

labor market in 1973 was a little less tight than in 1956. But the difference was not large enough to change my conclusion about the need for extreme caution, expecially since, as already noted, the critical, male-prime-age, labor market was tighter in 1972–73.

The interesting question at this point is whether the Council realized in late 1972 just how tight the situation was getting; the Report does indicate that the extraordinary rate of growth of the first quarter of 1973 was unexpected, and, since most other forecasters failed equally, I would not be prepared to blame the Council seriously for this failure. However, there is no indication that the tightness of the labor market was appropriately appreciated; quite the contrary, the January 1973 Report stated as its goal for 1973 to " . . . reduce the rate of unemployment to the neighborhood of $4\frac{1}{2}$ percent by the end of 1973 . . . " and that "This does not imply that in present circumstances $4\frac{1}{2}$ percent is necessarily the floor to the unemployment rate" (p. 73). These are amazing statements from the vantage point of the 1974 Report and one wonders why what was so obvious at the end of 1973 was apparently not even conceived of a year earlier, especially on the part of a Council that on the whole had been rather conservative in its unemployment targets.

To be sure, the 1974 Report stresses the fact that, from the beginning of 1973, fiscal policy stopped being highly stimulative. But neither did it become very restrictive. The budget shifted to a full-employment surplus of about \$3 billion by the conventional measure (Table 17) implying probably a budget balance on the Council preferred measure (inferred from a comparison of the two alternative measures of full-employment surplus for 1973, provided in Table 1, cols. 2 and 4). This measure, it should be acknowledged, however, tends to understate the restraining impact of fiscal policy, for the change from deficit to surplus was achieved primarily by holding outlays fairly constant while receipts were rising, and one must therefore allow for the "balanced budget multiplier" effect.

Looking next at monetary policy, in the first half of 1973 the money supply continued to rise at the hefty rate of nearly 8 percent. According to the data, the growth was rather erratic but I would not make much of this, especially since I mistrust the seasonal adjustment. Short-term interest rates rose fairly sharply by somewhat over 200 basis points; but long-term rates rose slowly and moderately, the new issue rate by 50 basis points and the mortgage rate by only 20, and both rates remained well below the peaks of the high unemployment period of mid-1971.

Unfortunately the 1974 Report's discussion of Monetary Policy in 1973, occupying but a page or so (pp. 83–84), is rather uninformative, and one cannot gauge whether, on the whole, the Council approved or disapproved the relatively expansionary policy being pursued, especially in the first half of the year. This is perhaps understandable in terms of the rather delicate relation and unclear division of power and responsibility between the Council and the Federal Reserve, a point on which I shall touch again later.

Finally in 1973 a further strong stimulus was provided to the economy by allowing, and even encouraging, the dollar to depreciate through the first half of the year. The associated swing in net exports added directly some \$3.5 billion to aggregate demand (presumably adding as much, to "offset to saving") in the first quarter, about one-tenth of the total increase in first quarter demand, and \$8.5 billion over the year, again close to 10 percent of the total increase in demand; and this quite aside from multiplier effects and direct effect on *U.S.* domestic prices, both

through imports and exports channels. The Report explicitly acknowledges these pervasive and, under the circumstances, undesirable effects of the depreciation on prices and aggregate demand. The effect on aggregate demand via net exports is attributed in part (pp. 55–56) to causes other than the depreciation, namely the boom and low food supplies in the rest of the world. At the same time, the list fails conspicuously to mention the effect of controls in Phase III and thereafter, such as the effect of price controls on exports and imports (although these too are acknowledged later in ch. 3's discussion of inflation control and in ch. 5, p. 190). Similarly the effect of devaluation on prices is explicitly recognized in chapter 3 (see p. 93). What is, however, conspicuously missing is an explanation and defense of the policy of letting the dollar continue to depreciate. There was a widespread feeling that the 10 percent official depreciation of February 12 was already somewhat on the high side. I shared this view with many other international trade specialists, such as Richard Cooper. But, as the Report recognizes (p. 186), there was an even more widespread feeling that the further depreciation in the second quarter by " . . . about 11 percent in terms of most of the EC currencies floating jointly and 5 percent in terms of the trade-weighted average of 14 currencies" (p. 185), (6 percent from March 19 to July, see pp. 92–93), was totally unwarranted and hence undesirable. I, for one, expressed this view both in print[4] and in private communications to the appropriate *U.S.* authorities. Nor have the events since provided any ground for changing that view.

The Report appears to foster the impression that this devaluation reflected the behavior of other central banks and not a *U.S.* game plan: " . . . when large-scale market intervention [by foreign central banks] failed to restore stability to foreign exchange markets, fixed exchange rates were abandoned; consequently the dollar fell . . . " (p. 183). But the simple truth is that the United States could have intervened on its own—and quite effectively as was shown by later experience in July—and that foreign Central Banks had stopped intervening when it became clear that the United States was not prepared to cooperate in an effort to support the dollar and, indeed, was not at all interested in the dollar being supported. On this point the Report states that in the third quarter " . . . there was limited intervention, by a number of central banks *including the United States, to prevent the dollar from rising*" (p. 183, italics added).

The peak of absurdity in our foreign exchange policy was reached in the third quarter of 1973 when we began to impose a variety of export restrictions, detailed in Table 24 of the Report, thus removing the only conceivable ground for the deep devaluation, namely to improve the current account balance. It would obviously have been far better from every point of view to push up the foreign value of the dollar and to impose export duties—as other countries have done—(whatever one finally concludes about the wisdom, from a longer run point of view, of limiting specific exports).

Chapter 2 also helps to make some sense of the apparent paradox mentioned in my opening paragraph, of a simultaneous sizable decline in real wages, reasonable growth in productivity, and yet, no substantial rise in profits and property income. The resolution of this puzzle actually involves a rather long chain, of which the Report covers only the last few links (pp. 74–75). In the first place, one finds that the share of labor in national income was essentially stable. The moderate rise in property income and the substantial rise

[4] In the *Corriere Della Sera* articles referred to earlier.

in farm income by nearly 45 percent in current dollars and 33 in terms of purchasing power—which, incidentally, was a good development from the point of view of reducing income inequalities—was offset by a substantial decline in the profits of nonfarm, noncorporate business, and professional incomes. The share was equally stable in terms of *GNP*, or Private *GNP*, and actually rose by about $1\frac{1}{2}$ percent in terms of Private Nonfarm *GNP*, after noncorporate business profits are corrected for imputed labor income. Hence real property income—outside farms—rose roughly as much as real output. But this rise *within* the year was only 4 percent (4.2 for private nonfarm) rather than 6 percent when measured year over year. As mentioned earlier, profits rose a little more and total property income a little less. But under these circumstances one would expect real wages to rise as much as productivity. Why, instead, did they decline significantly?

One part of the explanation is that productivity rose very little in the course of 1973. The increase in real output was achieved in fact by a quite sizable increase in employment—some 3.5 percent—and an almost equal increase in man-hours—3.4 percent for the private nonfarm economy—because hours worked apparently fell slightly. (However, the Report raises some reasonable doubts about the reliability of the hours worked series, p. 58.) Thus, measured productivity rose by less than 1 percent, the increase being concentrated in the first quarter, after which it tended to decline slightly. The second major explanation lies in the fact that even though hourly wages rose 6.7 percent, compensation per man-hour rose 8.2 percent (8 percent for private nonfarm). The difference between these two figures is accounted for, primarily, by the increase in "other benefits" (nearly 1 percent) " . . . the major part of which in 1973 came from the increase in employers' social security taxes in the first quarter" (p. 70), the rest being due to overtime and interindustry shifts. Since the private *GNP* deflator rose by just over 7 percent, real compensation per man-hour *measured in terms of the output produced* did rise by nearly 1 percent. For the nonfarm sector the increase was even larger, 2.5 percent, as the deflator rose by only 5.5, because of the 1.6 percent increase in the labor share. However, real wages are expressed in terms of consumables and not in terms of output produced. Now the price of consumption goods tended to rise relative to that of domestic output basically because of the sharp rise in import prices (26 percent), and even more relative to that of domestic nonfarm output (value-added) because of the surge in farm prices. Indeed, the consumption deflator rose by 7.4. However, the decline of real hourly wages was in terms of the cost of living index (*CPI*) which, as indicated, rose by 8.4. The difference between the two measures of the cost of consumables comes in part (about .5, p. 75) from the fact that the *CPI*, in contrast to the implicit deflator, uses fixed weights. The rest of the difference is due to the fact that the basket of goods used in the *CPI* is different from that in the implicit deflator; for example, it tends to give more weight to food which rose sharply.

The result of this long chain can be conveniently summarized by using the same links to explain another puzzle which is really the other side of the same coin: why did the *CPI* rise by as much as 8.4 percent when hourly wages rose only 6.7 and the profit share (the markup) actually declined? It should by now be apparent that the answer lies partly in the growing gap between hourly wages and the cost of labor to employers, through social security and other fringe benefits; this gap rose especially rapidly in 1973, the very year in which productivity grew very little. The

balance of the explanation is that the *CPI* rose much more than the implicit nonfarm deflator. This, in turn, can be accounted for in part by the sharply rising farm and import prices which caused a deterioration in the "terms of trade" for the domestic nonfarm sector; the remaining discrepancy between the *CPI* and the consumption deflator is due to the more questionable difference in composition between the basket of consumption goods actually consumed and the basket used in computing the *CPI*.

The moral of the story is, of course, that unless and until farm output expands and farm prices and import prices fall back, real labor income—including fringe benefits—must *pro-tantum* decline; and that a year in which productivity rises little is not the best suited for sizable increases in social security and other payroll taxes, at least if one cares for price stability.

The fact that all these unfavorable factors happened to hit the *U.S.* economy at the same time might well be considered a piece of bad luck for which the Administration could not be seriously blamed. It should also encourage some optimism about the future of inflation without the need to create a high rate of unemployment, just by avoiding a rapid expansion of demand when unemployment is already down to 5 percent. However, I am not quite prepared to go along with the first proposition about bad luck for the reason that at least two of the unfavorable factors—the high price of farm products and the high price of imports—can be attributed in some part to Administration action or inaction. A third factor, the poor performance of productivity, can also be at least partly attributed to an excessive expansionary policy in the presence of what should have been perceived as a relatively tight labor market and relatively tight productive capacity situation. I also suggest that it might well be attributed in part to the presence, and poor management, of price and related controls, a topic to which I now turn.

III. Did the Controls Help or Hinder—and How Much?

This issue is examined in chapter 3, which also gives a summary account of the major control measures and their administration. It was widely understood in the profession that the Council as well as Secretary Schultz had not been enthusiastic supporters of price controls—to say the least—a position which incidentally I also shared.[5] It was rumored, if only in jest, that its members were of two minds about its numerous failings since they vindicated their opposition and might also serve to immunize the public and Congress against the repetition of such arbitrary interferences with the system. This teach-you-a-lesson view certainly received support from the following characterization of price controls offered by Secretary Schultz in an interview to the Boston Globe: "It was a disaster from an economic view, but a great thing from an educational point of view." (!)

I was therefore very much intrigued by whether the Report would paint the episode as a success or as a failure. By and large it steers an intermediate or neutral course, at least in the summary conclusions: " . . . no one can disprove the thesis that the controls had a significant effect, although 1973 makes it a hard thesis to believe" (p. 108). These conclusions rely partly on an earlier section on "The Effectiveness of Controls"; this contains a theoretical disquisition, pp. 99–103, which I find rather muddy, the essence of which is that " . . . whether controls restrained the rate of inflation boils down to whether it can be demonstrated that they either restrained the rate of spending or increased the rate of production" (pp. 100–

[5] See my testimony before the Joint Committee cited earlier.

01). The effect on the rate of spending "... might show up in a lower ratio of spending and income to the money supply, or even in a lower money supply, if ... the Federal Reserve felt less need to permit monetary expansion" (p. 101). The main effect conceived on the supply side is of the classic type analyzed by Joan Robinson or in Abba Lerner's counterspeculation proposal: a ceiling on price by making marginal revenue coincide with price might raise marginal revenue and increase output. This framework might be a useful tool of analysis at the micro level but its macro relevance is rather questionable. Yet it is apparently on the basis of this type of analysis that the Report suggests that one cannot "... rule out the possibility that inflation might have been even greater in 1973 without controls. We think it could not have been much greater, however, since with the controls the rate of spending was high relative to the money supply, and output was low relative to the labor supply" (p. 108).

My own analysis of the effectiveness of controls, and I suspect that of many other colleagues, would have gone along very different lines. For some time now, I, as well as many others, have found that the behavior of aggregate wages and prices can be rather effectively analyzed in terms of a model both consistent in the theoretic sense and empirically useful. It consists of a generalized Phillips curve explaining hourly compensations in terms of unemployment, price expectations, and institutional variables (social security taxes, minimum wage legislation) and of a price equation in which the price level (and the real wage) is determined by an oligopolistic markup à la Sylos-Baine[6] on (long-run minimum) unit labor and raw material costs, with the markup stable in the relevant medium run, except for some upward and downward shading in response to the rate of utilization of capacity.[7]

Because of the lagged adjustment of wages to prices and prices to wages, this model implies that, once an inflationary process gets underway (say because of temporary excessive demand pressure on the labor market), it will take a rather long time for it to abate even if the rate of unemployment is kept substantially above the level consistent with reasonable long-run price stability. In terms of this structure, wage controls might be effective by holding down money wage increases, especially in the more highly unionized sector, partly by sheer decree and partly by reducing expectations of future price increases; and price controls might be effective partly by making wage controls acceptable to organized labor and partly by reducing the markup.

When the equations of such a model are estimated through a period ending in the third quarter of 1971, it is generally found that there is very little evidence of wage controls per se being effective in the sense that, given price behavior, there is no evidence of wages behaving significantly differently from the past. There is instead evidence of price controls being moderately effective in the sense that the markup on costs declines through the fourth quarter of 1973 (the end of Phase II) and then tends to move back toward the historical level, but by no means catching up with it at the end of 1973. The lower level of prices in turn has a feedback effect on wages, which feeds back on prices and so on, so that altogether one finds a non-negligible effect of Phase II on wages and a stronger one on prices, and similar effect

[6] See the author (1958).

[7] This structure, which is used in the MIT-PENN-Social Science Research Council (*MPS*) econometric model of the United States, is described in George de Menil and Jared Enzler.

on wages for Phase III, together with a more moderate one on prices.[8] On the basis of this evidence, obviously subject to many qualifications and uncertainties, I would be inclined to conclude, very tentatively, that Phase II was moderately successful and even Phase III helped, at least in the sense of slowing down the rebuilding of the markup.

However, these judgments, which at least for Phase II broadly agree with the Council's evaluation in the 1973 Report, relate only to the *aggregate* level of wages and prices, under *given* aggregate demand and productivity. To assess the overall effect we need to be concerned also with the effects of controls on these two other variables. Here it seems to me that the record must be assessed as rather negative in 1973. As the Report admits, price controls are likely to have stimulated exports, and hence aggregate demand pressure. The widely reported difficulty of procurement for many products cannot but have contributed to the very poor performance of productivity, something which the Report again acknowledges as likely on page 101. In addition, the unavailability of commodities to the consumers contributed of course to reduced welfare, even if this is not caught in our measure of real income and *GNP*.

The worst episode in the history of controls was clearly Phase III$\frac{1}{2}$ and particularly the notion that by freezing the prices of meats and poultry one could indirectly control the prices of feedstuffs—even though these were not directly controlled—because of the derived demand nature of the demand for these products. This maneuver could not possibly have worked given the fact that the inputs were storable, and prices were expected to be higher later, especially since the freeze was announced to last but two months. This policy led to the thinning out of supplies and contributed to the increase in food prices during August at the mammoth rate of some 80 percent at annual rates, an inference supported in part by the decline in food prices in the next two months following the end of the freeze.

On balance then one might agree, though on rather different grounds, with the Report's conclusion that the overall net effect of controls is hard to assess (except for Phase III$\frac{1}{2}$, which was definitely disastrous), but that on the whole, in 1973, given the erroneous aggregate demand management, the controls did appreciably more harm than good. One of the sad aspects of this conclusion is that it may have deprived us permanently of a device which, used extremely sparingly under conditions similar to those of 1972, might have been a useful addition to our almost empty box of tools to bring inflation to a rapid halt—without repeating the Great Depression!

IV. Targets and Policies for 1974—and their Consistency

Chapter 1 of the Report lays out a target path for real *GNP* for the year 1974, together with its implications for unemployment and prices, and with an outline of the fiscal and monetary policy designed to achieve that target. It is made clear that the figures for the first couple of quarters are more in the nature of a forecast than of desired rates because, given the lags in the response to policy, that initial period is largely beyond the reach of policy options, as of the time the Report was written. However, the path for the balance of the year is truly a "goal" as well as " . . . a prediction of what will be achieved if the planned policy is carried through" (p. 28). In this section I propose

[8] See, for instance, Robert Gordon, pp. 775–78, especially Table 1. The wage-price equations of the *MPS* model leads to very similar inferences.

to analyze the content of the chapter from three points of view: i) the appropriateness of the real output and employment targets; ii) their consistency with the associated price behavior; and iii) the consistency between the targets and the policies prescribed to achieve them.

The Report foresees for the first half "A slow rate of economic expansion . . . and possibly a decline, with rising unemployment" (p. 27). For the second half the target is essentially that of holding unemployment steady at the midyear level, by endeavoring to keep the growth of real *GNP* at roughly the long-run full-employment rate of 4 percent. This path implies a growth of real *GNP* of about 1 percent over 1973 and $1\frac{1}{2}$ percent within the year. No specific estimate is offered as to how high unemployment will have risen by midyear, but one can infer that it must be somewhere above 5.5 but short of 6, from the statement that "Unemployment for the year will be a little above $5\frac{1}{2}$ percent" (p. 28).

Relying on the forecasts of reliable analysts as well as on projection based on the *MPS* model I found myself in basic agreement with the assertion that policy could do little about the first half of the year, though I would like to know to what extent the rise in unemployment reflects the goals of policy pursued by the Administration in the second half of 1973, when economic conditions in the first half of 1974 were still within its control. The provisional first quarter figures, which in the meantime have become available, confirm that real *GNP* declined, and at a substantial rate, presumably larger than the Council expected and even larger than that of the most pessimistic forecasters.

As for the target in the second half, I share the view that it would be a mistake to try to reduce unemployment rapidly, but see little justification for going as far as holding it constant in the $5\frac{3}{4}$ range. I would much have preferred a goal of a modest reduction from the peak figure say down to $5\frac{1}{2}$ or just below, which, in my view, would have negligible cost in terms of inflation. This would require a rate of expansion closer to 5–6 percent on the average for the second half, bringing the year growth to 3–4 percent, but without raising very much the year-to-year growth. But these are relatively minor differences, especially if one has adequate humility about our ability to perform "fine tuning."

What is somewhat more serious is that, in my view, the Report is too optimistic about the behavior of prices consistent with the target real path; though opinions are divided on this point, my own model, as well as many others, suggests that the increase in the deflator year over year is likely to be closer to 8 percent than to the 7 percent projected by the Council. The first quarter figures suggest that even 8 may be optimistic, though I am not willing to be significantly swayed by the nearly 11 percent increase in the deflator reported for the first quarter. With this more realistic price forecast, the target path of money *GNP* should presumably involve a year-to-year growth just short of 9 percent rather than 8 percent (and a within-year growth of just about 9).

We now come to the most serious issue, that of appropriate policies. For fiscal policy the Report proposes a shift from a \$13.5 billion full-employment surplus to a \$1 billion deficit (based on the variable unemployment rate recommended by the Council). Together with this fiscal policy it recommends a monetary policy formulated—not surprisingly—in terms of "monetary aggregates" and more specifically in terms of M_2 (money plus time deposits except large CD's). M_2 is to increase at the rate of 8 percent on the ground that "For more than a decade the proportionate increase of money *GNP* tended to be the same as that of M_2, though in some years

the deviations from this proportionality were substantial . . . " (p. 32). The Report further suggests that this growth of M_2 would require a growth of M_1 of the order of 5 percent.

In cooperation with Wharton EFA, Inc. at the University of Pennsylvania, we have tried to simulate on the *MPS* model the effect of pursuing the Report's fiscal policy (as interpreted in the Wharton model), together with an 8 percent growth of M_2.[9] The basic result of the 8 percent growth of M_2 is that, beginning with the third quarter, the rate of growth of real *GNP* is *distinctly below* the Council projection: 2.5 percent in the third quarter and 3.4 in the last, with a within-year growth rate of only 1 percent, a year-to-year growth rate of about .7, fourth quarter *GNP* lower by $\frac{1}{2}$ percent and unemployment rising to 5.9. Furthermore the initial conditions for 1975 are seriously unfavorable; if the 6 percent growth of M_1 is continued into the first quarter, unemployment rises to 6.2, a figure which would not be significantly affected by an alternative growth of M_1 in that quarter, within reasonable limits.[10]

The reasons for these striking results are best understood by comparing them with the result of an alternative simulation in which monetary policy has been set so as to produce closely the real path targeted by the Council. It turns out that the required policy can be characterized as one of keeping the short-term interest rate—measured operationally by the 4–6-month prime commercial paper rate series—at around 8 percent, just a shade below the actual average level in the first quarter (8.3). The growth of M_1 required to achieve this result is about 8 percent within the year (which implies only 6.5 percent year over year, because of the flatness of M_1 in the second half of 1973). This implication for M_1 is hardly surprising considering that target money *GNP* rises by 9 percent and that, for a change, short rates are required to move slightly down within the year to achieve the stipulated real growth of *GNP*. The corresponding growth of M_2 is 10 percent within the year (as well as year over year), again reasonable under the circumstances.[11] The trouble with the Council recommended monetary policy is that the inadequate expansion of M_1 it implies perforce leads to sharply rising short-term rates: the implied commercial paper rate rises from 8.3 in the first quarter of 1974 to 9.5 in the last and 10.5 in the first quarter of 1975, the highest level in recent history. Similarly, the long-term corporate bond rate which, even under the "control" simulation, rises from the first quarter 7.9 to 8.1 would, with the Council's policy, rise to 8.6 by the first quarter 1975. These high rates would reduce real *GNP* by 1 percent ($9 billion) by early 1975 (residential structures by 10 percent, fixed investment by 1.3, consumption by .5).

The strong effect on residential construction comes partly from the fact that under present ceiling rates, the flow of savings into thrift institutions is adversely affected. This effect extends to time deposits and causes an 8 percent growth of M_2 to require a relatively larger (6 percent) growth of M_1 in the *MPS* simulation of the Council's policy. An alternative simulation of the Council's policy was therefore carried out, appropriately raising ceiling rates so as to maintain the recent historical relation between the growth of M_1 and M_2. We then find that an 8 percent growth of M_2 can be achieved with a growth of M_1 of only 5.5 percent. However, the results in

[9] The simulations were carried out by William Fitzgerald in consultation with Albert Ando. It might be noted that the 8 percent growth of M_2 turns out to require a growth of M_1 of around 6 rather than 5 percent partly for reasons mentioned below.

[10] For further details see the release by Wharton EFA, Inc., "MPS Model Forecasts, March, 1974—Experiment III."

[11] See the release cited in fn. 10.

terms of real *GNP* are essentially unaffected. The reason is that both short and long rates are roughly the same under either simulation of the Council's monetary policy for the reason that the rise in time deposit rates reduces the demand for money approximately as much as the reduction in supply.

It is of course entirely conceivable that the simulations from the *MPS* model are way off, and that the explanation suggesting their plausibility is equally wrong. Nonetheless I would urge the Council and the Federal Reserve to consider seriously the possibility that the achievements of the Council's own modest targets will require more growth of the monetary aggregates than recommended in the Report: in particular, the growth of M_2 should be closer to 10 than to 8 percent and that of M_1 closer to 8 than to 5. An 8 percent growth of M_1 was entirely too steep in the first half of 1973, given the moderate inherited rate of price rise, and in the face of unemployment that was low and falling, all of which created the need to raise promptly long-term rates to restrain an economy already overheated by past policies. But, in view of the present inherited high rate of inflation, the many nonrepetitive forces which have generated it, and the fact that unemployment is well on its way to pass the $5\frac{1}{2}$ percent mark by midyear, an 8 percent rate of growth of M_1 appears to me and to many others entirely appropriate to nudge down, not up, interest rates in order to support housing and other components of demand.[12]

What needs to be stressed is that the higher proposed rate of growth of M_1 and M_2 is no more than what is needed to achieve the Council's own modest target of a rate of unemployment above $5\frac{1}{2}$ percent throughout the second half of the year, and hence does not imply a higher rate of inflation than what is realistically consistent with that target.

My plea acquires a special urgency at this writing as the Federal Reserve appears to be terribly concerned over the fact that, since the beginning of the year, M_1 has grown at something like 8 percent and M_2 at something like 10, and is reacting by driving the Federal Funds rate above the 11 percent range, which in turn has already raised the commercial paper rate above 10 percent (instead of down to 8) and the corporate bond rate to the $8\frac{1}{4}$–$8\frac{1}{2}$ range. Does the Council and the Federal Reserve believe that these rates are consistent with Council's targets? Or has the Council abandoned even its modest targets? Or have they been abandoned, or even never accepted, by the Federal Reserve? The latter is perhaps the most disquieting possibility, since I hold strongly the view that it is for the Administration to set *GNP* targets (in cooperation of course with the Fed), and for the Fed to use its skills to achieve these targets—or persuade the Administration to change them—but not to aim at targets of its own.

V. Concluding Remarks

Because the previous sections are so critical of the Council's record, when I started on this manuscript I had hoped to devote some space to praise other sections of the Report and especially chapter 5 dealing with the distribution of income, which I have found a very useful and informative summary of past work, enriched by some results still unknown to me, and have enjoyed reading also in the light of my interest in the life cycle. Unfortunately the pressure to concentrate on urgent policy issues prevents me from saying any more. I do hope, however, that

[12] Almost identical conclusions have been set forth in a paper presented by Gordon in 1974, which the reader might find helpful since it provides a careful literary justification for the price forecast yielded by the *MPS* simulation, and a "monetarist-type" explanation of why a more expansionary policy is needed.

my criticism will be taken in the spirit in which it is offered—as an earnest endeavor to learn from past mistakes. And having been so candid in my criticism, I must be equally candid in acknowledging that I am glad indeed that it was not *my* responsibility to make decisions in the tough circumstances of last year and to write the Report, with Herbert Stein writing this review!

REFERENCES

G. de Menil and J. Enzler, "Prices and Wages in the FR-MIT-Penn Econometric Model," in *The Econometrics of Price Determination Conference*, Washington 1972.

R. Gordon, "The Response of Wages and Prices to the First Two Years of Controls," *Brookings Papers*, Washington 1973, *3*, 765–78.

———, "The Consequences of Inflation for Monetary Policy," presented at the Conference Board's Fifth Annual Midyear Outlook Conference, Chicago, Apr. 10, 1974.

A. P. Lerner, *The Economics of Control*, New York 1944.

F. Modigliani, "New Developments on the Oligopoly Front," *J. Polit. Econ.*, June 1958, *66*, 215–32.

———, *Testimony, Hearings before the Joint Economic Committee on the Economic Report*, 92d Congr. 1st sess., July 7, 8, 20, 21, 22, 23, 1971.

J. Robinson, *The Economics of Imperfect Competition*, London 1933.

U.S. Council of Economic Advisers, *Economic Report of the President*, Washington, Jan. 1973.

———, *Economic Report of the President*, Washington, Jan. 1974.

The 1974 Report of the President's Council of Economic Advisers: Energy in the Economic Report

By William D. Nordhaus*

It was the best of years and the worst of years. The Economic Report of the President proudly announces that the average income of Americans "reached record highs" (p. 4). Yet the index of consumer sentiment reached an all-time low in 1973.

It seems likely that the 1973 inflation was at the top of the list of economic ills, with consumer prices rising 8.8 percent over 1972. Most of the rise was due to severe structural shifts in agriculture and energy. In response to the unusual developments in those industries, chapter 4 of the Report is devoted to "Energy and Agriculture." I will concentrate only on the section on energy.

Most of the discussion of energy is a superficial narrative of recent history, with the usual tables of wholesale price data and pages filled with journalistic analysis. A sample of the style is the following history of petroleum usage:

> The use of petroleum products in the United States increased by 66 percent from 1960 to 1972. Much of this increase occurred in the transportation sector, which in 1972 accounted for 53 percent of the Nation's total petroleum use. A low excise tax has made the retail price of gasoline lower in the United States than in most other developed countries. The low gasoline price and a rapid growth in incomes have contributed to large increases in the number of motor vehicles on the road and in the total mileage driven, and thus to the rapidly growing demand for gasoline. Gasoline consumption has also been increased by the trend toward heavier automobiles with air conditioners and automatic transmissions, and by the use of emissions control devices. This expansion in demand for petroleum products was underestimated, as was the need for additional refinery capacity to meet that demand, with the result that the United States became heavily dependent on imports of refined products. [pp. 115–16]

It is interesting to ask how far off track naive predictions would have been. Since most of the shocks came after 1969, I ran a simple regression of total domestic demand for petroleum against time over the period 1947 to 1969. If we assume 1969 demand was on target, then according to this naive model, demand was 3.2 percent (500,000 barrels per day) percent above trend in 1972 and 4.6 percent (750,000 barrels per day) above trend in 1973. (A slightly more sophisticated regression including a *GNP* term shows a slightly larger error.) The magnitude of the surprise, then, could hardly have been larger than a 5 percent prediction error. Since this is well within the normal prediction range of a naive regression, and within the normal excess capacity of the petroleum industry, it is hard to see how these surprises by themselves could have caused a major dislocation.

I had looked forward to an enlightening discussion of the energy crisis. Again, I was disappointed. The section on The Energy Crisis contains no definition. In lieu of this, I propose the following definition of a "crisis": A market is in crisis

* Yale University.

when it shows *either* very sharp increases in price, or a quantitative restriction; in the latter case, of course, the effective price (as distinct from market or quoted price) shows a drastic increase.

By this definition, both agriculture and energy were crisis sectors in 1973. For food, there were no shortages (except during the ill-conceived second freeze), but wholesale prices rose by about 75 percent from November 1972 to August 1973. Perhaps the reason why the Report did not write about The Food Crisis was that there were no serious quantitative shortages.

For energy, the fall of 1973 and the winter of 1973–74 showed both price and quantity symptoms of crisis. By most accounts, the only shortfall was automotive gasoline, and here the "shortfall" ranged from 0 to 30 percent by state. The price consequences recorded in the Consumer Price Index were quite dramatic: From summer 1973 to February 1974, gasoline prices rose by 65 percent, home heating fuels by 55 percent, and electricity prices by 22 percent. Quantitatively, then, energy prices appeared to rise less than food prices. These are misleading, however, since there were quantity restrictions. It would be better to calculate the "demand shadow price," or the implicit price on the demand curve. (A colleague and I ran a small survey in New Haven and found that the shadow price was at a maximum (February) 14 cents a gallon or 28 percent further above the market, but by March this declined to 10 percent. These figures are roughly consistent with predictions which came from econometric studies of demand for gasoline.)

Whatever the exact dimensions, there is good evidence that the energy market was chaotic over the last half-year. What led to this problem? The Report explains the energy crisis as follows:

> The energy crisis originated in a large number of circumstances none of which was sufficient in itself to disrupt the economy seriously. Their convergence in 1972–73, however, touched off a dramatic change in the domestic energy supply-demand balance. [p. 117]

Four specific causes are mentioned: 1) "The natural gas price was [controlled] below the market-clearing price, . . . creating a shortage and leading to an increased demand for oil." 2) "The demand for oil was further increased by environmental restrictions . . . [and] delays in construction of nuclear reactors." 3) "Domestic refinery capacity was unable to meet the rapid expansion in demand for petroleum products." 4) "Although the domestic price of crude oil was supported [sic] above the price of imports, the price was not sufficiently high to discourage . . . an adequate expansion in domestic production" (p. 117).

In an elementary supply and demand framework, causes (1) and (2) led to an unexpectedly high rate of growth in demand. These seem basically correct, although (as I argued above) the quantitative magnitude of the surprise to a "naive" forecaster could hardly have been more than 5 percent by mid-1973.

Causes (3) and (4) are puzzling: (3) is a symptom of excess demand, not a cause; (4) works the wrong way; if price was supported above market clearing levels, why was there less supply than at the market clearing price? As we note below, keeping price high is one of the policies which would moderate the impact of a crisis.

Three other unmentioned factors seem equally important: First, the Arab oil embargo is implicit but crucial. This truly subtracted about 10 percent from petroleum supply. Second, the Administration's price controls are alleged to have prevented price reactions from clearing markets—this is certainly the case for the Simon regime of November 1973. These were probably also the reasons for the spot gaso-

line shortages in the summer of 1973. Finally, oil-import quotas—the original "Project Independence"—were extremely disruptive for planning purposes, leading to a lack of domestic refining capacity. With hindsight it is clear that the Administration's unwillingness to solve the problem of dependence in 1970, when the high-level review of the oil imports took place, put off the hard decision of supply reliability for four crucial years.

The Report next turns from discussion of the current crisis to the long-term prospects. This section is full of Polonial wisdom. Thus we are told:

> . . . large capital investments are required to expand domestic energy production. The private sector will be willing to undertake this investment only if there is a reasonable assurance that the price will remain sufficiently high to provide an adequate rate of return. [p. 122]

There are no answers to the tougher questions. In my mind, there are three important questions for policy makers in the current situation:

1. *Should the rate or direction of economic growth be changed because of impending global shortages of natural resources?* There is a serious school of thought holding that we are in for a sharp and continuing drop in per capita consumption standards; the line of argument is by now familiar to most economists. The view is given some superficial plausibility by the fact that the prices of crude materials rose 35 percent in 1973. Perhaps the Economic Report is not the place for such heavy thoughts, but I personally would have liked some insight into the Council's thinking.

2. *Are markets a reliable mechanism for allocating energy resources?* Although no explicit judgment is rendered, the Report implicitly assumed that free markets will do most of the allocation:

> These considerations argue for letting energy prices rise so that markets will clear, and for initiating a tax to limit windfall profits. In this way, the price system is permitted to play an important role in guiding production decisions and encouraging consumers to conserve energy. [p. 119]

In the past the petroleum industry has been heavily regulated, both by states via prorationing and by conservation regulations. Some of the regulations (such as well spacing) can be defended as a technique for overcoming the externalities of production. It has been increasingly clear that many of the regulations, especially those regulating natural gas prices, function quite perversely. The Report reiterates the Administration's proposal for deregulation of natural gas, and also argues for upward revision of the MER's for petroleum production. (The MER is defined as the maximum efficient rate of recovery. In principle, this is the rate which, if exceeded, would lead to avoidable loss of ultimate recovery; in practice, there is considerable doubt as to whether a moderate increase in the MER would lead to any loss.) Beyond these suggestions, there is no discussion of the more general question of the extent to which the market can be used for efficient allocation.

The reason for the reliance on markets as mechanisms for allocating resources is familiar to economists. Under fairly specific conditions—given preferences and technology and a complete set of futures and insurance markets—it can be shown that markets allocate resources efficiently. Unfortunately, a full set of markets is not available for most goods. For many reasons, markets are concentrated in the present, so only a very small fraction of energy resources are actually allocated over time by markets.

It can be seen that the main missing element in markets is the setting of the correct scarcity rent or shadow price on different resources. Without a full set of markets, the royalty is set mainly by the judgment of asset holders. How does this differ from other goods? Not much, except that the crucial royalty calculation is much more important for nonrenewable resources than for renewable resources, labor, or capital.

This brief discussion hints at the possible questions which can be raised about the role of the market in allocating energy resources. We can only wonder whether the Administration was also pondering these questions.

3. *What are the best ways of ensuring security of supply?* The discussion of "Project Independence" is one of the more interesting parts of this section. The Report is lukewarm about the idea of autarky, realizing full well the costs of an autarkic program. The Report throws out two suggestions for insulating the economy from insecurity of supply: storage and variable tariffs to bring import prices up to some guaranteed level. Beyond these casual suggestions, there is very little insight into the problem or the merit of alternative policies.

The problem of protecting our economic flank is terribly important. In what follows, I will analyze the suggestions put forward in the Report.

Consider a situation where a commodity such as petroleum is imported from an insecure source. The long-run supply curve is flat, with price at $\bar{p}$ for "normal" years, but the supply is subject to interruptions. The interruptions are characterized by a decrease in supply to a fraction μ for a period of time, and this interruption is judged to have probability π. For the United States, petroleum is a natural example of an insecure source, where the insecurity arises from political embargoes. For other countries, food is probably a similar problem but the insecurities are intrinsic given variations in the weather.

One policy for "treating" the problem of insecurity is to raise the cost of foreign supply so as to reduce "dependency" on foreign source. The dependency takes the form of greater inelasticity of domestic supply and demand in the short run rather than in the long run.

To be specific, we assume the long-run demand for imports is

$$q_0 = a - b(\bar{p} + t) = a - bp_0 \tag{1}$$

where p is price including the tariff, t is the tariff on the insecure foreign source, and q is net demand for imports. The short-run demand curve has elasticity λ times the long-run elasticity $(0<\lambda\leq 1)$. If p_1 and q_1 are the post-embargo price and quantity, this implies that

$$(p_1 - p_0) = -\frac{1}{b\lambda}(q_1 - q_0) \tag{2}$$

Consider the gains and losses to a tariff; in all that follows, the tariff can be interpreted as that tariff above and beyond the "classical" arguments for tariffs, such as infant industry, optimal tariff, etc. In normal years, the tariff will raise the cost to domestic consumers, causing a welfare loss equal to $\frac{1}{2}t\Delta q$, or equal to the area of the shaded triangle in Figure 1.

In "embargo" years, however, the reduced dependency lowers the social cost of reduced supply. This is illustrated in Figure 2. Regime I has a low tax and leaves the economy with high imports and on short-run demand curve SR^I. A percentage reduction of supply will lead to deadweight loss of shaded area $ABCD$. (The loss is calculated from the viewpoint of the importing country, not the world.) A high tariff in regime II on short-run demand curve SR^{II} will lead to lower initial consumption at q_0^{II}, and thus to less absolute loss in

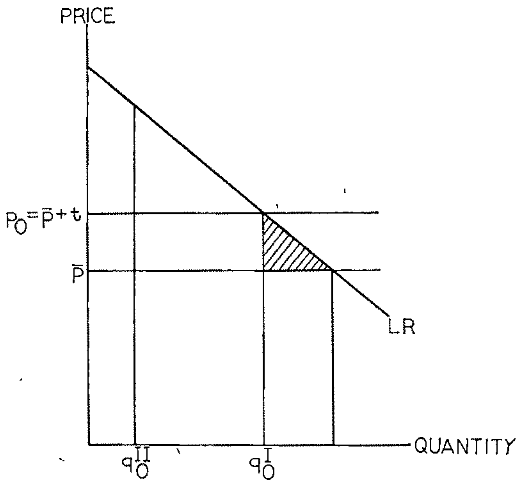


FIGURE 1

shaded area *EFGH*. In the limit, as complete autarky is reached, there is no cost to a supply interruption.

With the simplifying assumptions given above, we can calculate the optimal tariff as follows. Measuring a loss as a negative number, the loss from the tariff (L_T) is:

$$L_T = -\tfrac{1}{2}t^2 b \tag{3}$$

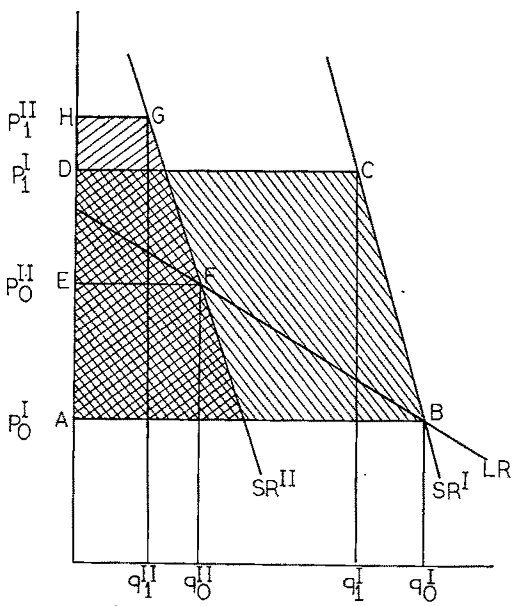


FIGURE 2

The loss from the embargo (L_E) is:

$$L_E = -(p_1 - p_0)(q_1 + q_0)/2 \tag{4}$$

Since $q_1 = \mu q_0$ and from (2), $p_1 - p_0 = -(q_1 - q_0)/b\lambda = -q_0(\mu - 1)/b\lambda$, we have

$$L_E = \frac{q_0(\mu - 1)(1 + \mu)q_0}{2b\lambda}$$

and from (1): $$L_E = \frac{[a - b(\bar{p} + t)]^2(\mu^2 - 1)}{2b\lambda}$$

The criterion function is then $L(t)$, showing loss as a function of the tariff:

$$L(t) = L_T + \pi L_E$$

or

$$L(t) = -\tfrac{1}{2}t^2 b + \phi[a - b(\bar{p} + t)]^2 \tag{5}$$

where $$\phi = \frac{\pi(\mu^2 - 1)}{2b\lambda} < 0$$

We desire to minimize the loss. This implies

$$L'(t) = 0 = -tb - 2\phi b[a - b(\bar{p} + t)]$$

or $$t = -2\phi[a - b(\bar{p} + t)]$$

or

$$t = \frac{2\phi(a - b\bar{p})}{2\phi b - 1} \tag{6}$$

To get some idea of the magnitude of the optimal tariff, assume that interruptions occur on average every fifth year, and that the short-run elasticity is five times as inelastic as the long-run elasticity. An embargo is assumed to reduce imports by one-quarter. Further assume that the long-run demand curve has a zero-import level at $\bar{p} + t = 12.5$ and imports 6 million barrels per day at $p = 5$.[1] Considering the optimal tariff with $\bar{p} = 1.5$ (pre-1973) and $\bar{p} = 7$ (post-1973), we then get:

[1] Thus $Q = 10 - 0.8\, p$ and $\phi = 0.2 \times (.75^2 - 1)/2 \times .2 \times .8 = -.27$.

	Imported Price	Optimal Tariff	ϕ	Domestic Price	q_0
Case 1	\$1.50	\$3.32	−0.27	\$4.82	6.14
Case 2	\$7.00	\$1.66	−0.27	\$8.66	3.07

Note that we can explicitly evaluate the loss function. For $\bar{p}=1.5$, we have losses as follows:

t	L_T	πL_E	L
0	0.00	20.91	20.91
$t^*=3.32$	4.41	10.19	14.60
$t=8.5$	28.90	0	28.90

The optimal policy is roughly one-half as costly as autarky and two-thirds as costly as free trade. The model also shows why Project Independence is poor economics: it always pays to take a little risk if the odds are in your favor.

In a world of certainty, an import quota and a tariff can be equivalent. The presumption in favor of tariffs is that the revenue goes to the government rather than to firms. In a world of uncertainty, the problem is much more complicated and involves the relative uncertainties about supply and demand curves. Note, however, that the optimal tariff or quota is quite restrictive. For Case 1, the tariff is twice the foreign price.

A tariff-quota arrangement is only one of a number of possible methods for preventing overdependence on foreign sources. It has the virtue that both foreign and domestic oil are priced at their proper values (in terms of consumption). There are, however, other policies which are worth examining. Most other policies in effect reduce the insecurity by having inventories to smooth out the transition between the short- and long-run situations.

Storage performs the function of being an alternative source of supply. Storage will pay if it covers cost. If storage cost is ν per year, then storage pays if $\nu/\pi<(p_1-\bar{p})$ in the optimal arrangement above. Thus in Case 1, price during embargo years is

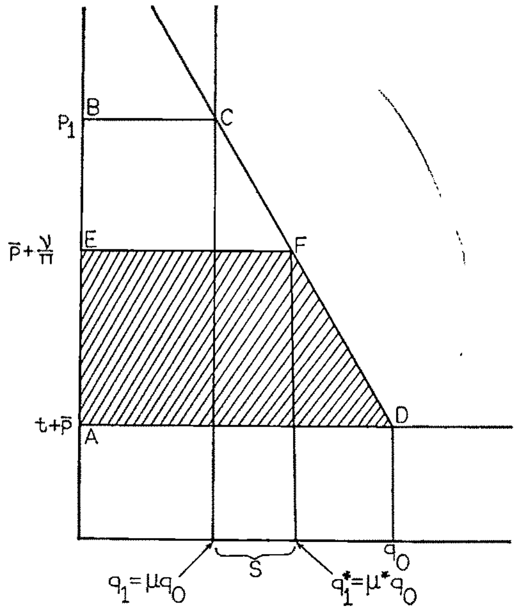


FIGURE 3

14.41, so if storage costs are less than 2.58, it pays to meet at least part of the demand by storage.

More generally, note that a given tariff induces a short-run import demand curve such as that shown in Figure 3. Let us assume that a cutback to μq_0 occurs, giving a temporary price of p_1. It will pay to store amount S to bring the price during embargo years back down to $\bar{p}+\nu/\pi$.

We can short-circuit the rather complex analysis by noting that the storage policy completely neutralizes any effect on price beyond $\bar{p}+\nu/\pi$. The optimal policy will therefore be to calculate the effective μ^* (determined by optimal storage), then set the tariff *as if* the embargo were at level μ^*. Thus in Figure 3, μ^* is determined by the point to which storage is taken.

To calculate μ^*, note that from (2)

$$\bar{p}+\frac{\nu}{\pi}=p_1=p_0-\frac{1}{b\lambda}(q_1-q_0)$$

$$=p_0-\frac{1}{b\lambda}(\mu^*-1)q_0$$

(over)

$$p_1 - p_0 = -\frac{(\mu^* - 1)}{b\lambda}[a - bp_0]$$

$$\text{so } (\mu^* - 1)(a - bp_0) = b\lambda\left(\bar{p} + t - \bar{p} - \frac{\nu}{\pi}\right)$$

$$(\mu^* - 1) = \frac{b\lambda\left(t - \frac{\nu}{\pi}\right)}{a - b(\bar{p} + t)} \tag{7}$$

Substituting (7) and (6) we get the new loss function, L^*. To obtain the optimal storage-tariff program, solve (6) and (7) simultaneously. Assume ν is 0.5. An approximate solution using the same data given above is:

	Import Price	μ^*	t^*	Domestic Price	q_0	Storage
Case 1	\$1.5	.968	.65	\$2.15	8.28	1.81
Case 2	\$7.0	.938	.59	\$7.59	3.93	0.74

A storage-tariff arrangement thus will generally have a lower tariff and some storage as a substitute. In the case outlined above, the tariff is slightly above 0.5, while storage represents between 15 and 20 percent of imports times the expected length of the embargo. For Case 1, in embargo years price would rise from 2.15 to 4.00, reducing quantity by about 5 percent. In any case, the tariff will always be less than ν/π.

The total cost of the storage-tariff package is as follows:

	L_T	πL_E	νS	Total
Case 1	−0.17	−2.70	−0.90	3.8
Case 2	−0.14	−1.16	−0.37	1.7

The storage option reduces the total loss considerably. Comparing the loss in the storage-tariff option with the tariff alone, we see loss is reduced by 75 percent.

Reserve capacity serves as another form of storage. The question of whether reserve capacity or storage from foreign sources is better is a simple cost calculation. If the landed price of foreign crude plus storage costs is less than the marginal cost of developing capacity, then the storage scheme is preferable; and vice versa.

Using the principles laid out above, how should future energy policy be set? Taking the figures used above as suggestive, the following policies would be consistent with the analysis presented above:

1. As a first approximation, set the tariff on foreign-produced crude oil at about \$0.60 per barrel. To implement this, however, a tax on consumption of \$0.60 will be levied, with a \$0.60 rebate to domestic producers. Thus, storage of foreign oil will be at cost, and domestic oil will be consumed at cost; foreign oil will be consumed at cost plus the tariff. (It should be stressed that the \$0.60 figures comes out of the parameters assumed above. Different assumptions would lead to different figures.)

2. In order to give incentives for secure supply, the \$0.60 tax could also be rebated to those countries (Canada? Venezuela?) whose governments are willing to ensure supply for a period of time.

3. Once the tax has been levied, a storage program would be implemented so that a target level of *reserve* storage is achieved. The target level might be in the order of one year's imports, depending on the best estimates of demand and supply. Storage could be bonded as reserve, with release only by special designation. To induce the storage, special storage certificates could be sent out for bids. If the storage policy has been properly set, these should cost the government very little, since they are designed to break even.

4. If the above policies are deemed insufficient to protect national defense or other crucial industries, those sectors should make explicit estimates of their own needs and have their own stockpiles.

The costs of these stockpiles can be explicitly charged to these sectors rather than having them spread diffusely and inefficiently over the entire economy.

5. Since proposals 1 to 4 are deemed adequate to guarantee supply and protect the national interest, all other measures designed to foster artificially the extraction of domestic energy resources (except new technologies) should be abolished. We should therefore abolish the depletion allowance, current write-off of intangibles, treating royalties as income taxes; obviously, oil import quotas should not be reinstated.

This seems to be the logic of the Economic Report. Having spent all this space analyzing the rationale for tariffs or storage, it must be asked why markets will not perform the function adequately. Note that the model used above assumes that there is a one-way option—output will be reduced but never increased. Put differently, the expected value of output $E(q)$ is $[(1-\pi)\cdot 1+\pi\cdot\mu]\cdot q_0<q_0$; whereas consumers and producers choose techniques as if $E(q)=q_0$. If consumers are rational, with μ having positive variance and mean of 1, the answer is quite different. In this case, $E(\phi)>0$, so the optimal regime in (6) is for a subsidy rather than a tariff on the unstable import!

If a good set of futures markets for petroleum existed, then much of the storage function could be performed by speculators. As long as speculators are not too risk averse they will store enough so that $p_1=\bar{p}+\nu/\pi$, exactly the efficient condition. The problem is serious, however, because there are no broad futures markets for petroleum in the United States. It might be argued that development of such markets would be a better way of ensuring an efficient hedge against supply interruptions than would the tariff-storage proposal described above. Until such markets exist, however, it seems that the government has no alternative to ensuring sufficient storage supply in a direct manner.

Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis

By P. A. SAMUELSON AND S. SWAMY*

Index numbers of prices and quantities, an invention of the last century, are important for macro-economic description and for sliding-scale wage and other contracts. Desirable properties for an index of many items have been postulated by Irving Fisher (1922) and others by analogy with the case of a single good's price or quantity. If a single good's price doubles, the index should double; the index between any two dates will not be changed if the base period of the index is changed from one date to another; a dimensional change in the good (as from grams to pounds) should not change the index, nor should a dimensional change in money (as from pennies to dollars or dollars to pounds); finally, for a single good, the product of the index number of price and the index number of quantity will equal the ratio of the values of total expenditure between any two compared dates.

Although Ragnar Frisch (1930) has proved that, when the number of goods exceeds unity, it is impossible to find well-behaved formulae that satisfy *all* of these Fisher criteria, we derive here canonical index numbers of price and quantity that do meet the spirit of all of Fisher's criteria in the only case in which a single index number of the price of cost of living makes economic sense—namely, the ("homothetic") case of unitary income elasticities in which at all levels of living the calculated price change is the same. This seeming contradiction with Frisch is possible because the price and quantity variables are not here allowed to be arbitrary independent variables, but rather are constrained to satisfy the observable demand functions which optimize well-being.

The invariant quantity and price index numbers are shown to be "dual" in exactly the same sense that the production function and the minimum-unit cost function are dual; further, the price index function is shown to be the reciprocal of the indirect-utility dual function in its canonical, homogeneous first-degree form. By contrast, in the nonhomothetic demand case, the central role of the base quantity level of living in the price index is shown to have as a natural dual a central place for a base price level in an index of quantity expressed in terms of standardized dollars. Economic theory warns that no single index-number formula could correctly characterize alternative preference and indifference-contour patterns. Brief treatment is also given here to taste changes, to index-number approximation, to path-dependent and path-independent Divisia line integrals, and to index-number measures of changes in production possibilities; new ground is broken, and a fairly complete survey of the theory of index numbers is sketched. Some of our work

* Massachusetts Institute of Technology, and Indian Parliament, New Delhi, respectively. We owe gratitude to our common teacher Wassily Leontief. Samuelson acknowledges financial aid from the National Science Foundation that enabled Jill Pappas to give us valuable editorial assistance. We have benefitted from conversations or correspondence with Robert Pollak, Robert Summers, Dale Jorgenson, Hajime Hori, and Meir Kohn.

overlaps with the independent findings of Robert Pollak and Sydney Afriat, and of the now classic work of Ronald Shephard.

I. Economic Assumptions

One set of ordinal preferences is assumed to be observed involving n goods $(q_1, \ldots, q_n)=Q$. (Throughout we use capital letters for row vectors and lowercase or Greek letters for scalars.) Preferences can be summarized by any indicator of ordinal preference $u(Q)$ or $f\{u(Q)\}$ where $f'\{\ \}>0$. The prices of the goods are $(p_1, \ldots, p_n)=P$. Expenditure on the goods is denoted by $e=p_1q_1+ \ldots +p_nq_n=P\cdot Q=Q\cdot P$. Superscripts are used to denote the dates (or places) of the observations as in (Q^1, P^1), (Q^0, P^0), (q_j^1), (p_j^0), etc.

Historically,[1] one can discern three different approaches to index-number theory.

1) Jevons, Edgeworth, and other early writers tended to think of them as some kind of a mean or measure of central tendency of a universe of price changes over time. (Thus, if each price ratio could be regarded as a random item drawn independently from the same skew distribution, which is approximately lognormal, then the geometric mean would be the maximum likelihood statistic that most accurately estimates its parameter of location.)

2) Later, as with Irving Fisher, Walsh, Palgrave, Persons, Allyn Young, Frickey, and other writers of the World War I epoch, certain mechanical tests were applied to index-number formulas.[2] From the two-plus-two-equals-four tautology that the geometric mean of an array of positive numbers lies between the arithmetic and harmonic means, it was somehow thought that the arithmetic mean had *"upward bias."* Exactly what zero bias meant was never thought through.

3) Finally, attention returned toward the economic theory of a price index[3] (associated with older names of Marshall, Lexis, Wicksell, . . . and more modern names of Könus, Pigou, Haberler, Keynes, Bortkiewicz, Gini, Bowley, H. Schultz, Staehle, Frisch, Leontief, Lerner, R. G. D. Allen, Wald, Ulmer, Wold, Samuelson, Theil, Afriat, Gilbert and Kravis, Beckerman, Pollak, Fisher and Shell, . . .). This paper confines itself to this economic approach, with its following definitions.

DEFINITION: *Economic Price Index: This must equal the ratio of the (minimum) costs of a given level of living in two price situations.*

Not quite so obvious as the price case is the corresponding definition for a quantity index.

DEFINITION: *Economic Quantity Index: This measures for two presented quantity situations Q^0 and Q^1, the ratio of the minimum expenditure needed, in the face of a reference price situation P^α, to buy their respective levels of well-being.*

Although most attention in the literature is devoted to price indexes, when you analyze the use to which price indexes are generally put, you realize that quantity indexes are actually most important. Once somehow estimated, price indexes are in fact used, if at all, primarily to "deflate" nominal or monetary totals in order to arrive at estimates of underlying "real

[1] The best general references are still the 1936 surveys of Frisch, Leontief, and Hans Staehle (1935). See also Samuelson (1947, ch. 6), which is in the spirit of Leontief. Note too our earlier citation of Pollak and Afriat. See also the useful survey by Richard Ruggles.

[2] See Fisher (1911, 1922) and Frisch (1930).

[3] See Melville Ulmer for the following quotation: "Writing in 1707, William Fleetwood, Bishop of Ely, set himself the task of determining the relative difference in money income which would provide for a student of the University of Oxford 'the same Ease and Favour' in his day and 260 years before" (p. 28). The fifteenth century B.C. Indian treatise *Arthastra* shows similar concerns.

magnitudes" (which is to say, *quantity* indexes!).

We begin by sketching the price case. Write the required index in the equivalent forms

$$(1.1)\quad p[P^1, P^0 \mid u(Q^a)] \equiv p(P^1, P^0; Q^a)$$

By its definition, the price index can be written as

$$(1.2)\quad \begin{aligned} p(P^1, P^0; Q^a) &= e(P^1; Q^a)/e(P^0; Q^a) \\ &= e[P^1 \mid u(Q^a)]/e[P^0 \mid u(Q^a)] \end{aligned}$$

where by definition

$$(1.3)\quad \begin{aligned} e(P; Q^a) &\equiv e[P \mid u(Q^a)] \\ &= \operatorname*{Min}_{Q} P\cdot Q \end{aligned}$$

such that $\quad u(Q)=u(Q^a)$

It follows from their definition that these e functions are homogeneous of degree one and concave in the prices; and for regular, reversible demand curves where the indifference contours are strongly convex from below, e will also be a quasi-concave function of Q.

We now describe mathematically the economic quantity index.[4] By its definition, we can write this in the following notation:

$$(1.4)\quad q(Q^1, Q^0; P^\alpha) = e(P^\alpha; Q^1)/e(P^\alpha, Q^0)$$

Here P^α is the arbitrarily chosen price standard for which the quantity comparison is to be made.

The fundamental point about an economic quantity index, which is too little stressed by writers, Leontief and Afriat being exceptions, is that it must itself be a cardinal indicator of ordinal utility. That is, $q(Q, Q^0; P^\alpha)$ must, for Q^0 and P^α fixed, be itself of the form $f\{u(Q)\}$. Likewise $q(Q^1, Q; P^\alpha)$ must, for Q^1 and P^α fixed, be another version of $-f\{u(Q)\}$. These facts are guaranteed by our economic definition. But the point is worth stressing because it shows at the beginning that we cannot hope for one ideal formula for the index number: if it works for the tastes of Jack Spratt, it won't work for his wife's tastes; if, say, a Cobb-Douglas function can be found that works for him with one set of parameters and for her with another set, their daughter will in general require a non-Cobb-Douglas formula! Just as there is an uncountable infinity of different indifference contours—there is no counting tastes—there is an uncountable infinity of different index-number formulas, which dooms Fisher's search for *the* ideal one. It does not exist even in Plato's heaven.[5]

The fundamental and well-known theorem for the existence of a price index that is invariant under change in level of living Q^a, is that each dollar of income be spent in the same way by rich or poor, with all income elasticities exactly unity (the homothetic case). Otherwise, a price change in luxuries could affect only the price index of the rich while leaving that of the poor relatively unchanged. This basic theorem was well known already in the 1930's, but is often forgotten and is repeatedly being rediscovered.

HOMOGENEITY PRICE THEOREM: *If, and only if, preferences are homothetic, with all income elasticities unitary, and with one $f\{u(Q)\}$ being of the canonical form $q(Q)\equiv\lambda^{-1}q(\lambda Q)$, a homogeneous first-degree function, will there exist an invariant index,*

[4] For a somewhat different definition of the quantity index in the general nonhomothetic case, see Pollak, pp. 60–61. In terms of later sections' dual "indirect utility function," $u^*(P/e)$, our general price and quantity indexes can be described thus: solve $u^*(P^\alpha/e)=u(Q^0)$ and $u^*(P^\alpha/eq)=u(Q^1)$ for $q=q(Q^1, Q^0; P^\alpha)$, the quantity index; and solve $u(Q^a)=u^*(P^0/e)=u^*(P^1/ep)$ for $p=p(P^1, P^0; Q^a)$, the price index.

[5] This point, forcefully made in Samuelson (1947, p. 154) has served to intimidate many (but, alas, not all) searchers for the nonexistent. See for example Doris Iklé, on which we comment later.

$$(1.5)\quad p(P^1, P^0; Q^a) \equiv p(P^1, P^0; Q^b) \equiv p(P^1, P^0) = p(P^1)/p(P^0)$$

for all (Q^a, Q^b)

Proof for this will be reserved for the next section.

On reflection, it will also be clear that in the homothetic case, and only then, will the quantity index be the same for *any* chosen price standard P^α.

HOMOGENEITY QUANTITY THEOREM: *If, and only if, preferences are homothetic with all income elasticities unitary with* $q(Q)=\lambda^{-1}q(\lambda Q)$ *being an admissible cardinal indicator of utility, will for all* P,

$$(1.6)\quad q(Q^1, Q^0; P^\alpha) \equiv q(Q^1, Q^0; P) \equiv q(Q^1, Q^0) = q(Q^1)/q(Q^0)$$

Proof of this theorem will be reserved for the next sections, which will also relate the invariant indexes of price and quantity to 1) the duality theory of production and cost, elaborated by R. W. Shephard, Samuelson (1953), Hirofumi Uzawa, and Daniel McFadden; 2) the duality theory of direct and indirect utility, associated with Harold Hotelling, René Roy, Hendrik Houthakker (1960), Samuelson (1965), and Pollak; 3) the interrelations between the two dualities, as in Samuelson (1972); and 4) certain new aspects of duality.

II. Duality of Prices and Quantity Indexes

The following theorem establishes the exact canonical form of the invariant $p(P^1, P^0)$ index.

FIRST DUALITY THEOREM: *Ratios of the homogeneous first-degree function* $p(P)\equiv\lambda^{-1}p(\lambda P)$ *and* $q(Q)\equiv\lambda^{-1}q(\lambda Q)$, *provide respectively, the invariant canonical price and quantity indexes, namely*

$$(2.1)\quad p(P^1, P^0) = p(P^1)/p(P^0)$$
$$q(Q^1, Q^0) = q(Q^1)/q(Q^0)$$

where $q(Q)$ has been defined as a homogeneous first-degree numerical indicator of homothetic preferences, unique up to an arbitrary scale or dimensionality constant, and where $p(P)$ is its dual defined by

$$(2.2)\quad p(P) = \min_Q P\cdot Q/q(Q)$$

$$p(P/e)q(Q) \leq 1$$

with equality holding only for optimal $(Q,\ P/e)$ demand pairing $Q=D(P/e)$, $P/e=D^{-1}(Q)$, and where $p(P/e)$ is concave, homogeneous first-degree in its arguments. In the regular case when $u(Q)$ is strongly quasiconcave, the $q(Q)$ function will have these same properties as its dual $p(P/e)$, and the dual to the dual of $q(Q)$ will be $q(Q)$ itself. The reader may verify the dual counterpart to (2.2)

$$(2.2')\quad q(Q) = \min_P Q\cdot P/p(P)$$

and, also, that the contour, $p(p_1/p_q, \dots) = 1$, defines the useful "factor-price frontier," p_q being the price of the q output. We take these duality relations as established by cited works of Shephard, Samuelson, Afriat, and others.

We now proceed to prove the two homogeneity theorems. We begin with sufficiency of homothetic tastes for the independence of the price index from the chosen level of living. Provided $q(Q) = \lambda^{-1}q(\lambda Q)$, the homothetic case, $P\cdot Q/q(Q) \equiv P\cdot\lambda Q/q(\lambda Q)$ and takes on for fixed P and $q(Q)=q(Q^a)$ all the values that it takes on for the same fixed P and no restrictions on the Q. Hence, by its definition in (1.3) and setting $\lambda=q(Q^i)^{-1}$,

$$(2.3)\quad e(P; Q^i) = \min_Q P\cdot Q$$

such that $q(Q) = q(Q^i)$

$$= q(Q^i) \min_Q P\cdot Q/q(Q) =$$

$$= q(Q^i)p(P),$$
from the definition of (2.2)

Therefore

$$(2.4)\quad \begin{aligned} p(P^1, P^0; Q^i) &= e(P^1; Q^i)/e(P^0; Q^i) \\ &= q(Q^i)p(P^1)/q(Q^i)p(P^0) \\ &= p(P^1)/p(P^0) \\ q(Q^1, Q^0; P^\alpha) &= q(Q^1)p(P^\alpha)/q(Q^0)p(P^\alpha) \\ &= q(Q^1)/(Q^0) \end{aligned}$$

which proves sufficiency of homotheticity.

To prove necessity of homotheticity, note that the functional equation

$$(2.5)\qquad \frac{e(P^1; Q^i)}{e(P^0; Q^i)} \equiv \frac{e(P^1; Q^j)}{e(P^0; Q^j)}$$

implies for the monotone $e(;)$ function that it be factorable into

$$(2.6)\qquad e(P; Q) = a(P)b(Q)$$

where we know that $a(P)=\lambda^{-1}a(\lambda P)$ because e always is homogeneous first-degree in its P arguments. Shephard and Afriat have noted (2.6)'s factorization theorem for the homothetic case, and valid for it only.

The invariance of the price index is seen to imply and to be implied by the invariance of the quantity index from its reference price base

$$(2.7)\quad q(Q^1, Q^0; P^\alpha) = \frac{e(P^\alpha; Q^1)}{e(P^\alpha; Q^0)} \equiv \frac{e(P^\beta; Q^1)}{e(P^\beta; Q^0)} = \frac{a(P^\alpha)}{a(P^\alpha)}\,\frac{b(Q^1)}{b(Q^0)}$$

To keep our necessity argument brief, we revert to a calculus proof. From the classical formulation of Lionel McKenzie or the standard envelope theorem of maximization subject to a parameter, we know

$$(2.8)\quad \partial e(P; Q)/\partial p_i = q_i \qquad (i=1, \dots, n)$$

$$(2.9)\quad \begin{aligned} \frac{\partial e/\partial p_i}{\partial e/\partial p_j} &= \frac{\partial\{a(P)b(Q)\}/\partial p_i}{\partial\{a(P)b(Q)\}/\partial p_j}, \qquad \text{from (2.6)} \\ &= q_i/q_j = \{\partial a(P)/\partial p_i\}/\{\partial a(P)/\partial p_i\} \\ &= \text{function of } (p_2/p_1, \dots, p_n/p_1) \text{ alone} \end{aligned}$$

But this last relation is evidently the case only when homotheticity prevails. Hence, necessity of homotheticity as well as sufficiency is proved.

III. Indirect Utility Duality and Dual Dualities

In this compressed survey, we note briefly the relationship between the $[p(P),\ q(Q)]$ duality and the indirect-direct utility duality $[u(Q),\ u^*(P/e)]$ where

$$(3.1)\qquad u^*(P/e) = \operatorname*{Max}_{Q} u(Q)$$

such that $Q\cdot P/e = 1$

Note that $u^*(P/e)$ differs in sign from the full duality of Samuelson (1965). When $u(Q)\equiv q(Q)\equiv\lambda^{-1}q(\lambda Q)$ and $P/e = D^{-1}(Q)$, it is easily shown as in (2.3) and Samuelson (1972), that

$$(3.2)\quad \begin{aligned} u^*(P/e) &= eu^*(P) = u(Q) \\ u(Q)p(P) &= u(Q)[u^*(P)]^{-1} = e, \qquad \text{from (2.2) and (3.2)} \\ p(P) &= 1/u^*(P) \end{aligned}$$

Hence

$$(3.3)\qquad p(P^1, P^0) = u^*(P^0)/u^*(P^1)$$

This leads to the unification theorem of the last-cited paper.

DOUBLE DUALITY THEOREM: *For $u(Q)\equiv q(Q)$, the "production dual" $p(P/e)$ equals the reciprocal of the "indirect-utility" dual, $u^*(P/e)$.*

Some further aspects of duality may be mentioned. Swamy has found it convenient to utilize a related duality that treats a numeraire good asymmetrically. Solve for

q_1, the expression

$$u(q_1, \ldots, q_n) = u(Q^a) \tag{3.4}$$

to define a new $m(q_2, \ldots, q_n; Q^a)$ primal function.

$$\begin{aligned} -q_1 &= m(q_2, \ldots, q_n; Q^a) \\ &= m[q_2, \ldots, q_n \mid u(Q^a)] \end{aligned} \tag{3.5}$$

$$\partial m(q_2, \ldots; Q^a)/\partial q_i = p_i/p_1 \qquad (i = 2, \ldots, n) \tag{3.6}$$

Then it is seen that $e(P; Q^a)/p_1$ is dual to $m(q_2, \ldots; Q^a)$ in the sense that

$$\begin{aligned} &e(P; Q^a)/p_1 \\ &\quad = m^*(p_2/p_1, \ldots, p_n/p_1; Q^a) \\ &\quad = \operatorname*{Min}_{Q} \sum_1^n p_j q_j / p_1 \end{aligned} \tag{3.7}$$

such that $q_1 + m(q_2, \ldots; Q^a) = 0$

$$\begin{aligned} e(P; Q^a)/p_1 = \operatorname*{Min}_{q_{i+1}} \Big[&-m(q_2, \ldots, q_n; Q^a) \\ &+ \sum_2^n (p_j/p_1) q_j \Big] \end{aligned}$$

$$\partial m^*(p_2/p_1, \ldots, p_n/p_1; Q^a)/\partial(p_i/p_1) = q_i \qquad (i = 2, \ldots, n) \tag{3.8}$$

This Hicks-Allen marginal rate of substitution parent function $m(\ )$ and its dual $m^*(\ )$ hold for the general nonhomothetic case. However, in the special homothetic case where some $u(Q) = q(Q) \equiv \lambda^{-1} q(\lambda Q)$

$$\begin{aligned} &q_1 + m(q_2, \ldots; Q^a) \\ &\qquad \underset{\lambda}{\equiv} \lambda q_1 + m(\lambda q_2, \ldots; \lambda Q^a) \end{aligned} \tag{3.9}$$

$$\begin{aligned} &m[q_2, \ldots, q_n \mid q(Q)] \\ &\qquad \equiv \lambda^{-1} m[\lambda q_2, \ldots, \lambda q_n \mid \lambda q(Q)] \\ &q_1/q^* = -m[q_2/q^*, \ldots, q_n/q^* \mid 1] \end{aligned}$$

where q^* is a parameter taking on any scale we wish and the function on the right is the solution for q_1/q^* of the implicit-function equation

$$1 = q(q_1/q^*, q_2/q^*, \ldots, q_n/q^*) \tag{3.10}$$

IV. Passing Fisher's Tests: The Homothetic Case

We can now show that *all* of the traditional test criteria of Fisher (1911) for an index number are satisfied by the canonical pair $[p(P^1, P^0), q(Q^1, Q^0)]$ in the homothetic case.

COMPLETENESS THEOREM: *These canonical index numbers satisfy, for any number of goods $n \geq 1$, the Fisher test criteria appropriate to the primitive $n=1$ case.*

$$\text{(i) If } P^1 = \lambda P^0,\ p(P^1, P^0) = p(\lambda P^0, P^0) \equiv \lambda, \tag{4.1}$$

"general mean of price relatives"

$$\text{(ii) } p(P^1, P^0)\, p(P^0, P^1) \equiv 1, \tag{4.2}$$

"time-reversal test"

$$\text{(iii) } p(P^2, P^1)\, p(P^1, P^0) \equiv p(P^2, P^0), \tag{4.3}$$

"circular-reversal test"

Properties (i)-(iii) are seen to be *also* exactly satisfied for $q(Q^1, Q^0)$, where q and p, and Q^i and P^i, are interchanged respectively.

(iv) Dimensional change from q_j to $q_j\dagger = d_j q_j$, $q\dagger = d_{n+1} q$, $e\dagger = d_{n+2} e$, $d_j > 0$, leave indexes invariant[6]

[6] The literature, from Fisher on, including Samuelson (1967, p. 25) and Swamy (1965, p. 620), is inadequate on the dimensional invariance test. Properly speaking, once one has introduced the appropriate dimensional constants, we impose thereby no restrictions on the functional form of the index number. Thus, write in our notation the Frisch-Fisher index functions of $4n$ rather than $2n$ variables, namely $f(P^1, Q^1; P^0, Q^0)$. Now subject all q's to arbitrary dimensional changes $q_j\dagger = d_j q_j$, $d_j > 0$. We must now have $f\dagger(P\dagger^1, Q\dagger^1; P\dagger^0, Q\dagger^0)$ give identical results as $f(P^1, Q^1; P^0, Q^0)$. But note that $f\dagger(\ \ )$ is a *new and different function* from $f(\ \ )$. As an example, consider $f = p(P^1)/(p(P^0) = (p_1^1 + p_2^1)/(p_1^0 + p_2^0)$, which seems to fail the (ill-conceived) test. Actually, as we can see from the case $(d_1, d_2) = (2, 1)$, it leads to the proper $f\dagger = (p_1\dagger^1 d_1^{-1} + p_2\dagger^2 d_2^{-1})/(p_1\dagger^0 d_1^{-1} p +_2 \dagger^0 d_2^{-1}) = (\frac{1}{2} p_1\dagger^1 + p_2\dagger^1)/(\frac{1}{2} p_1\dagger^0 + p_2\dagger^0) \equiv f$. See Percy W. Bridgman for proper treatment of dimensional analysis in the natural sciences and logic. (Samuelson, it should be said, no longer regards his brief discussion in (1967, pp. 24–25) as fully optimal; it is superseded by the present paper.) Further

$$(4.4)\qquad p\dagger(P\dagger^1, P\dagger^0) \equiv p(P^1, P^0)$$
$$q\dagger(Q\dagger^1, Q\dagger^0) \equiv q(Q^1, Q^0)$$

"dimensional test"

$$(4.5)\qquad \text{(v)}\ \ p(P^1, P^0)q(Q^1, Q^0) \equiv e^1/e^0$$

"factor-reversal[7] test"

Property (i) follows immediately from the first-degree homogeneity property of $p(P)$ as defined in (1.6). Property (iii) follows immediately from the ratio form of $p(P^1, P^0)$ as $p(P^1)/p(P^0)$ and it implies (ii) as a special binary case, as well as implying the general case

$$(4.6)\qquad p(P^n, P^{n-1})p(P^{n-1}, P^{n-2}) \dots p(P^1, P^0) \equiv p(P^n, P^0)$$

Property (iv) follows from

$$\begin{aligned} q(Q) &= q(q_1, \dots, q_n) \\ &= q(q_1\dagger d_1^{-1}, \dots, q_n\dagger d_n^{-1}) \\ &= d_{n+1}^{-1} q\dagger(q_1\dagger, \dots, q_n\dagger) \\ &= d_{n+1}^{-1} q\dagger(Q\dagger) \end{aligned}$$

$$\begin{aligned} P\cdot Q/q(Q) &= \sum_1^n p_j q_j [q(Q)]^{-1} \\ &= d_{n+2}^{-1} \sum_1^n p_j\dagger d_j d_j^{-1} q_j\dagger d_{n+1} \\ &\quad \cdot [q\dagger(Q\dagger)]^{-1} \\ &= d_{n+2}^{-1} d_{n+1} P\dagger\cdot Q\dagger/q\dagger(Q\dagger) \end{aligned}$$

$$p(P) = d_{n+2}^{-1} d_{n+1} p\dagger(P\dagger)$$

$$(4.7)\qquad p\dagger(P\dagger^1, P\dagger^0) = \frac{d_{n+1}^{-1} p(P^1)}{d_{n+1}^{-1} p(P^0)} = p(P^1, P^0)$$

$$(4.8)\qquad q\dagger(Q\dagger^1, Q\dagger^0) = \frac{d_{n+1} q(Q^1)}{d_{n+1} q(Q^0)} = q(Q^1, Q^0)$$

Property (v)'s factor-reversal test is seen to be satisfied from the original definition of the quantity index, combined with (2.3)'s definition of $p(P)$ and the identities

$$(4.9)\qquad p(P^1, P^0)q(Q^1, Q^0) = p(P^1)q(Q^1)/p(P^0)q(Q^0) = e^1/e^0$$

This completes the proof. It remains only to note that, for $n=1$, $p(P^1, P^0) = p_1^1/p_1^0$, $q(Q^1, Q^0) = q^1/q^0$ and all of Fisher's properties (i)–(v) follow trivially.

How have we shown to be possible what Frisch (1930) proved to be impossible? Of course, that is not quite what we have done.

First, and least essential, Frisch followed the old practice of adding a regularity condition that we have not postulated. It is the so-called "determinateness test," which requires that, as some $p_j \to 0$ or ∞, the index should not go to 0 or infinity. This condition, it seems to us, is an odd one and not at all a desirable one. Thus, this Frisch regularity condition is *necessarily* violated in the trivial one-good case where $n=1$, and it rules out the nonsatiation assumptions often made in standard economic theory.[8]

discussion of the Frisch-Fisher axioms is provided in Samuelson (1974a).

[7] Since we do not impose the strong requirement that the *same* formula apply to p as to q, for reasons to be discussed later, our condition (v) might well be called the "weak factor-reversal test," in contrast to the more common "strong" case where $p(X, Y) \equiv f(X, Y) \equiv q(X, Y)$. Nonetheless, as indicated in Samuelson (1965) and Swamy (1970), the Cobb-Douglas does provide one singular case where $p(P)$ and $q(Q)$ are "strongly self-dual," namely the Cobb-Douglas case where each has the form $f(X) = n^{\frac{1}{2}} \prod_1^n x_i^{k_j}$, $\sum_1^n k_j = 1$. Samuelson has recently discovered an infinity of other self-dual cases; and Wahidul Haque of Toronto seems to have made similar, as yet unpublished, discoveries. The unwary may be tempted by a fallacious way of getting non-Cobb-Douglas invariant indexes that will satisfy the strong rather than weak factor-reversal tests. Let $q(Q)$ and $p(P/e) = p(Y)$ be dual functions of different form. Why not mate them to get a new pair of the form $\{[q(Q)p(Q)]^{1/2}, [q(Y)p(Y)]^{1/2}\}$? Alas, these last are not dual to each other. Even though

$$q(Q)p(Y) \leq 1,\ q(Y)p(Q) \leq 1$$
$$[q(Q)p(Y)]^{1/2}[q(Y)p(Y)]^{1/2} \leq 1$$

the equality signs will *not* hold simultaneously for the optimal-demand pairings $Q = D(Y)$, $Y = D^{-1}(Q)$, unless $D(X) \equiv D^{-1}(X)$, the strongly self-dual case that holds from the beginning, with no need for mating!

[8] For critical discussions on Frisch's 1930 treatment

Second, writing in 1930 before the economic theory of index numbers had, thanks to Frisch's own work and that of others, displaced the mechanical earlier formulations, Frisch had no need to insist upon the demand-dependence holding between $Q=(q_1, \ldots, q_n)$ and $P/e=(p_1/e, \ldots, p_n/e)$. His index formulas are postulated from the beginning to be functions of the arbitrary [!] $4n$ variables (P^1, Q^1, P^0, Q^0), rather than of the $2n$ variables (Q^1, Q^0) *or* the $2n$ variables $(P^1/e^1, P^0/e^0)$.

In any case, it is easy for us to specify instances of existent well-behaved and admissible price and quantity index numbers that satisfy *all* the tests (i)–(v), including the strong form of (v)'s factor-reversal test in which $p(,)$ and $q(,)$ are identical functions.

Thus, let the expenditures on each good be constant positive fractions of total expenditure $(k_1=1-\sum_2^n k_j, k_2, \ldots, k_n)$, the Cobb-Douglas case.[9] Then

$$(4.10)\qquad \begin{aligned} q(Q) &= n^{\frac{1}{2}} \prod_1^n q_j^{k_j} \\ p(P) &= n^{\frac{1}{2}} \prod_1^n p_j^{k_j} \\ q(Q^1, Q^0) &= \prod_1^n (q_j^1/q_j^0)^{k_j} \\ &= \prod_1^n (q_j\dagger^1/q_j\dagger^0)^{k_j} \\ p(P^1, P^0) &= \prod_1^n (p_j^1/p_j^0)^{k_j} \\ &= \prod_1^n (p_j\dagger^1/p_j\dagger^0)^{k_j} \\ p(P^1, P^0)q(Q^1, Q^0) &= \prod_1^n (k_je^1/k_je^0)^{k_j} = e^1/e^0 \\ &= \prod_1^n [(p_j^1/p_j^0)(q_j^1/q_j^0)]^{k_j} \end{aligned}$$

The difference between our procedure, in which either of the $2n$ variables (q_j^1, q_j^0) or $(p_j^1/e^1, p_j^9/e^0)$ determines the other set, is brought out by examining the last two lines in (4.10).

Obviously,

$$(4.11)\qquad \prod_1^n \left(\frac{p_j^1 q_j^1}{p^0 q^0}\right)^{k_j} = \frac{\sum_1^n p_j^1 q_j^1}{\sum_1^n p_j^0 q_j^0}$$

is *not* an identity in the $4n$ indicated variables *all* regarded as *arbitrary*. But it *is* an identity satisfied by all values of the $4n$ variables that also satisfy (4.10)'s optimality demand relations

$$(4.12)\qquad p_j^i q_j^i = k_j e^i \quad (j = 1, \ldots, n; i = 0, 1)$$

We can now answer this question: What is the widest class of admissible invariant index numbers $[q(Q^1, Q^0), p(P^1, P^0)]$? Ob-

of this determinateness point, see the references in Swamy (1965) to Subramanian, Mizutani, and Wald. Fisher made little of this requirement in his 1922 book, merely carrying it over from his 1911 classic. As mentioned, to insist on this test requirement would be to rule out the nonsatiation assumptions often made in economic theory. Thus, if $q=(2q_1q_2)^{1/2}$, lowering any one price enough should yield bliss beyond that of any preassigned finite increment of income; and why rule this out? For both $q(Q)$ and $p(P)$ to satisfy this strict criterion would by itself and without reference to other test criteria entail a self-contradiction. We do not deny sympathy with Frisch's practical point that "... the withdrawal or entry of any [new] commodity will often have to be performed as a limiting case when either the quantity ... or the money value ... decreases toward zero, respectively increases from zero" (1930, p. 405). Thus, $q=\sum a_j q_j$, $p=\text{Min}[\ldots, p_j/a_j, \ldots]$ would handle this kind of case where one new gadget adds only a bit to well-being. But we do not wish to impose *our* wishes on the economic agents who can have any tastes they wish—including cases where some good (potassium, say) is vitally necessary in minimum ratios, or alternative cases of nonsatiation. Therefore, with good conscience we do not impose the self-contradictory determinateness test.

[9] Actually, for the homothetic quadratic-utility function the Ideal index is exact, which is the Buscheguennce-Alexander case mentioned in Samuelson (1947) and in Afriat. In view of the discussion of (4.18), this suggests that, along the $Q=D(Y)$ function of the quadratic-homogeneous case, the Ideal index does satisfy the circular test after all. Of course, the Ideal index satisfies only the weak form of (v).

$$(4.13a) \qquad q = \left[\sum_1^n a_j q_j^\gamma \right]^{1/\gamma}, \quad 0 \neq \gamma < 1, \qquad \text{"CES"}$$

$$p = \left[\sum_1^n p_j^\delta / a_j \right]^{1/\delta}, \qquad (1-\gamma)(1-\delta) = 1$$

$$(4.13b) \qquad q = \sum_1^n a_j q_j, \qquad \text{"Laspeyres-Paasche quantity index"}$$

$$p = \text{Min}\,[p_1/a_2, \ldots, p_n/a_n]$$

$$(4.13c) \qquad q = \text{Min}\,[q_1/a_1, \ldots, q_n/a_n], \qquad \text{"fixed proportion case"}$$

$$p = \sum_1^n a_j p_j, \qquad \text{"Laspeyres-Paasche price index"}$$

$$\lim \gamma \to 1 \text{ of (4.13a) } CES \text{ case}$$

$$(4.13d) \qquad q = n^{1/2}(a_1 q_1)^{k_1}(a_2 q_2)^{k_2} \ldots (a_n q_n)^{kn}$$

$$1 - \sum_1^n k_j = 0,\ k_j > 0,\ \lim \gamma \to 0 \text{ of (4.13a), the Cobb-Douglas case}$$

$$p = n^{1/2}(p_1/a_1)^{k_1}(p_2/a_2)^{k_2} \ldots (p_n/a_n)^{k_n}$$

$$(4.13e) \qquad q(Q) = q_1 \sum_1^n [\mathit{log}\,(a_j q_j/q_1) - \exp\,(b_j q_j/q_1)]$$

---

viously, there are as many as there are arbitrary homogeneous first-degree functions $q(Q) \equiv \lambda^{-1} q(\lambda Q)$ or $p(P) \equiv \lambda^{-1} p(\lambda P)$. Thus, we can generate new homothetic preference functions and invariant indexes *ad lib*. The examples (4.13a)–(4.13e) illustrate that there are uncountably many more "exact" index-number formulas than those few previously recognized: the Cobb-Douglas geometric-mean case; the so-called Ideal index; etc.

We must give a brief but unavoidably technical digression to explain why there does not exist, for *any* old-fashioned index-number formula arbitrarily preassigned, some indifference contours for which the formula is exact. Fisher (1922) gives literally thousands of different formulas of the general form

$$(4.14) \quad f(Q^1, P^1; Q^0, P^0) \equiv g(Q^1, Y^1; Q^0, Y^0), \quad Y \equiv P/e$$

$$g(Q^0, Y^0; Q^1, Y^1) g(Q^1, Y^1; Q^0, Y^0) \equiv 1$$

Since (4.14) follows the usual presupposition that the price and quantity $g(\ )$ function are to be "of the same form," the next section's discussion of test (i^*) will show that only in the homothetic case will the generalized mean property hold for $q(\lambda Q^0, Q^0; P^\alpha) \equiv \lambda$ as well as for $p(\lambda P^0, P^0; Q^\alpha) \equiv \lambda$. So we may examine (4.14)'s possibilities under the simplifying assumption of homotheticity. It is well known for this case that the relations hold that appear as (6.3) and (6.3′) below, and which we now borrow for use here

$$(6.3) \quad p_i/e = \partial\, \mathit{log}\, q(q_1, \ldots, q_n)/\partial q_i \qquad (i = 1, \ldots, n)$$

$$Y = \mathit{grad\ log}\ q(Q)$$

Hence, you might at first think that setting

$$(4.15) \quad q(Q) \equiv g(Q, \mathit{grad\ log}\ q(Q); Q^0, Y^0) q(Q^0)$$

would, in the homothetic case, yield a first-order partial differential equation

that, given the form of $g(\ )$, could be solved for a corresponding $q(Q)$, in terms of whose indifference contours the $g(\ )$ formula would be exact. However, for this to yield a consistent result, it is necessary that $g(\ )$ have certain special properties. Thus

$$(4.16)\quad q(Q) \equiv \lambda^{-1}q(\lambda Q) \quad \text{implies}$$
$$g(\lambda Q, \lambda^{-1}Y; Q^0, Y^0) \equiv \lambda g(Q, Y; Q^0, Y^0)$$

Also, consider *all* Q^0 on a base indifference curve passing through an arbitrarily selected point Q^*, with $q(Q^0)=q(Q^*)=1$. In terms of (3.5), $[(Q^0, Y^0)]$ can then be written as

$$(4.17)\quad \overset{0}{q_1} = -\, m(\overset{0}{q_2}, \dots, \overset{0}{q_n}; \overset{*}{Q})$$

$$(4.17')\quad \overset{0}{y_1} = \left[-m(\overset{0}{q_2}, \dots, \overset{0}{q_n}; \overset{*}{Q}) + \sum_{2}^{n} \overset{0}{q_j}\partial m/\partial \overset{0}{q_j} \right]^{-1}$$

$$(4.17'')\quad \overset{0}{y_i} = \overset{0}{y_1}\partial m/\partial \overset{0}{q_i} \quad (i = 2, \dots, n)$$

Hence, we can replace the $2n$ parameters (Q^0, Y^0) in $g(\ )$ by $n-1$ arbitrary $q_2^0, \dots, q_n^0$ parameters along the base indifference contour. Now (4.15) becomes (after we substitute (4.17), (4.17′), (4.17″) into it):

$$(4.18)\ q(Q) = h[Q, \textit{grad log } q(Q); \overset{0}{q_2}, \dots, \overset{0}{q_n}]$$
$$\equiv \lambda^{-1}h[\lambda Q, \lambda^{-1}\, \textit{grad log } q(Q); \overset{0}{q_2}, \dots, \overset{0}{q_n}]$$
$$q(\overset{*}{Q}) \equiv h[\overset{*}{Q}, \textit{grad log } q(\overset{*}{Q}); \overset{0}{q_2}, \dots, \overset{0}{q_n}]$$

If an arbitrary Fisher or new formula of type given in (4.14) is to be exact for some preference field, the solution to (4.18) must be independent of $q_2^0, \dots q_n^0$! And why, in general, should general $f(\ )$, $g(\ )$, $h(\ )$ formulas have this needed property? As an example, we should be surprised if the Iklé proposed formula has this property; we do not even know whether its quantity index is unity on all Q^1 points for which $u(Q^1)=u(Q^0)$.

The last part of this section is basic to deep insight into the meaning of the five Fisher-Frisch tests in the general nonhomothetic case to which we now turn. For the pathologies considered in (4.18)'s solution are as nothing compared to the pathologies to be expected in the general nonhomothetic case.

V. Commentary on Fisher-Frisch Tests in the General Nonhomothetic Case

We have already spoken on the gratuitousness of the determinateness test that requires an index not to go to zero or infinity as a subset of its variables go to these limits; and also of the erroneous interpretation of the requirement of dimensional invariance. We must stress again that the factor reversal test offers no stumbling block for our definitions of $p(P^1, P^0; Q^a)$ and $q(Q^1, Q^0; P^a)$ if, as we should do logically, we drop the *strong* requirement that the *same* formula should apply to $q(Q)$ as to $p(P)$. A man and wife should be properly matched; but that does not mean I should marry my identical twin!

Where most of the older writers balk, however, is at the circular test that frees us from one base year. Indeed, so enamoured did Fisher become with his so-called Ideal index

$$(5.1)\quad p(P^1, P^0)$$
$$= [(P^1 \cdot Q^0/P^0 \cdot Q^0)(P^1 \cdot Q^1/P^0 \cdot Q^1)]^{1/2}$$
$$= \text{square root of (Laspeyres} \times \text{Paasche)}$$
$$= (\lambda_p \pi_p)^{1/2}$$

that, when he discovered it failed the circular test, he had the hubris to declare "..., therefore, a *perfect* fulfillment of this so-called circular test should really be taken as proof that the formula which fulfills it is erroneous" (1922, p. 271). Alas, Homer has nodded; or, more accurately, a great scholar has been detoured on a trip whose purpose was obscure from the beginning.

If all the other tests were satisfied, but the circular test were violated and if $q(Q, Q^0; P^\alpha)$ is to be one cardinal indicator of utility, we would be facing a contradiction. Thus, suppose we are given the data

$$(5.2) \quad q(Q^2, Q^0; P^\alpha) = q(Q^1, Q^0; P^\alpha)$$

From the circular test we could infer

$$(5.3) \quad q(Q^2, Q^1; P^\alpha) = 1$$

and from it

$$(5.3') \qquad u(Q^2) = u(Q^1)$$

The last is a proper inference. Yet if (5.2) did not imply (5.3) it could not imply (5.3′)! Also the following contradictions would be possible if the circular test could be violated.

$$(5.4) \quad \begin{aligned} q(Q^2, Q^0; P^\alpha) &= q(Q^1, Q^0; P^\alpha) = 1 \\ q(Q^3, Q^0; P^\alpha) &= q(Q^1, Q^0; P^\alpha) = 1 \\ q(Q^3, Q^2; P^\alpha) &> 1 \end{aligned}$$

This leads to the absurd intransitivity

$$(5.5) \qquad u(Q^2) = u(Q^1)$$

$$\text{and} \qquad u(Q^3) = u(Q^1),$$

$$\text{but} \qquad u(Q^3) > u(Q^2)$$

Conclusion: So long as we stick to the economic theory of index numbers, the circular test is as required as is the property of transitivity itself. And this regardless of homotheticity or nonhomotheticity.

Furthermore, as will be shown in our next section on approximation, Fisher missed the point made in Samuelson (1947, p. 151) that knowledge of a third situation can add information relevant to the comparison of two given situations. Thus Fisher contemplates Georgia, Egypt, and Norway, in which the last two each have the same price index relative to Georgia:

> We might conclude, since "two things equal to the same thing are equal to each other," that, therefore, the price levels of Egypt and Norway must be equal, and this would be the case if we compare Egypt and Norway *via* Georgia. But, evidently, if we are intent on getting the very best comparison between Norway, we shall not gō to Georgia for our weights . . . [which are], so to speak, none of Georgia's business. [1922, p. 272]

This simply throws away the transitivity of indifference and has been led astray by Fisher's unwarranted belief that only fixed-weights lead to the circular test's being satisfied (an assertion contradicted by our $p(P^1)/p(P^0)$ and $q(Q^1)/q(Q^0)$ forms).

We may now ask the question, "In the nonhomothetic case, how do the index numbers fare relative to the Fisher test criteria?" That his five tests cannot then be successfully passed is seen by verifying the following:

$$(5.6) \qquad \text{(i*)} \quad \begin{aligned} p(\lambda P^0, P^0; Q^\alpha) &\equiv \lambda \\ q(\lambda Q^0, Q^0; P^\alpha) &\not\equiv \lambda \end{aligned}$$

Thus, as our quantity index is defined, even the first of Fisher's tests fails in the nonhomothetic case, as any example will reveal. A simple instance is where

$$(5.7) \qquad u(Q) = \text{Min}\,[q_1 + 1, q_2]$$

$$Q^0 = (2, 3), \quad Q^1 = \lambda Q^0, \quad \lambda > 1$$

$$q(\lambda Q^0, Q^0; P^\alpha) = \frac{(3\lambda - 1)p_1^\alpha + 3\lambda p_2^\alpha}{2p_1^\alpha + 3p_2^\alpha}$$

$$= \lambda - \frac{(\lambda - 1)p_1^\alpha}{2p_1^\alpha + 3p_2^\alpha} < \lambda$$

If, like Pollàk, one employs a quantity definition that satisfies Fisher's (i*), then one of the other tests, such as (v*), will fail in the nonhomothetic case—and this even though the $(Q^1, P^1; Q^0, P^0)$ are not treated as $4n$ arbitrary independent variables, but rather do respect the equilibrium demand relations

$$(5.8) \qquad Q^i = D(P^i/P^i \cdot Q^i) \qquad (i = 0, 1)$$

It is easy to show that our defined pairs $p(P^1, P^0; Q^0)$ and $q(Q^1, Q^0; P^1)$, or $p(P^1, P^0; Q^1)$ and $q(Q^1, Q^0; P^0)$, satisfy *all* the other Fisher tests (ii–v), (v) being understood as the "weak" factor-reversal test of footnote 7. But before showing this, it is well to elucidate that, in the nonhomothetic case, there is no real symmetry or full-duality between the definitions of price and of quantity indexes. One sees this from the original definitions, from examining footnote 4's mathematical asymmetric formulations of the definitions in terms of the genuinely dual variables $(P/e, Q) \equiv (P/P\cdot Q, Q) \equiv (Y, Q)$. A formulation in terms of the nondual variables $(P, Q) \equiv (Ye, Q)$, as in $p(P^1, P^0; Q^a)$ and $q(Q^1, Q^0; P^\alpha)$, could not be expected to be completely dual. Put differently, we can match (5.6)'s asymmetric property in which $q(\ ,\ ;)$ fails a test in the nonhomothetic case by the following asymmetric formulation in which $p(\ ,\ ;)$ fails a test—or, more fairly, lacks a property the $g(\ ,\ ;)$ enjoys. Thus, by basic demand theory with its lack of money illusion, it never matters for the quantity index if we multiply all reference prices P^α by a common factor λ, and use λP^α. Thus, by contrast with (5.6), we have

$$(5.9)\quad \begin{aligned} q(Q^1, Q^0; \lambda P^\alpha) &\equiv q(Q^1, Q^0; P^\alpha) \\ p(P^1, P^0; \lambda Q^a) &\not\equiv p(P^1, P^0; Q^a) \end{aligned}$$

If Fisher had adjoined to (i*) the requirement that the quantity index is never to be affected by scale changes in P^1 or P^0 (which leave their "weightings" unchanged), we'd learn in the nonhomothetic case that both indexes must fail this widened (i*) test. It is only from the fact that in the homothetic case, the references used, P^α or Q^a, never matter (cancelling out, so to speak), that (5.6) and (5.9) get satisfied in the homothetic case.

We now speedily show the sense in which the tests (ii*)–(v*) can all be satisfied even in the nonhomothetic case.

$$(5.10)\quad \text{(ii*)}\ \begin{aligned} p(P^1, P^0; Q^a)p(P^0, P^1, Q^a) &\equiv 1 \\ &= \frac{e(P^1; Q^a)e(P^0; Q^a)}{e(P^0; Q^a)e(P^1; Q^a)} \\ q(Q^1, Q^0; P^\alpha)q(Q^0; Q^1; P^\alpha) &\equiv 1 \\ &= \frac{e(Q^1; P^\alpha)}{e(Q^0; P^\alpha)}\,\frac{e(Q^0; P^\alpha)}{e(Q^1; P^\alpha)} \end{aligned}$$

$$(5.11)\quad \text{(iii*)}\ \begin{aligned} &p(P^2, P^1; Q^a)p(P^1, P^0; Q^a) \\ &\quad\cdot p(P^0, P^2; Q^a) \equiv 1 \\ &q(Q^2, Q^1; P^\alpha)q(Q^1, Q^0; P^\alpha) \\ &\quad\cdot q(Q^0, Q^2; P^\alpha) \equiv 1, \end{aligned}$$

by similar proofs;

$$(5.12)\quad \text{(iv*)}\ \begin{aligned} &p\dagger(P\dagger^1, P\dagger^0, Q\dagger^a) \\ &\quad\equiv p(P^1, P^0; Q^a) \\ &q\dagger(Q\dagger^1, Q\dagger^0; P\dagger^\alpha) \\ &\quad\equiv q(Q^1, Q^0; P^\alpha), \end{aligned}$$

by the same earlier dimensional analysis.

When it comes to the factor-reversal test, we easily verify the "crossed" condition

$$(5.13)\quad \text{(v*)}\ \begin{aligned} &p(P^1, P^0; Q^1)q(Q^1, Q^0; P^0) \\ &\quad\equiv e^1/e^0 \\ &\quad\equiv p(P^1, P^0; Q^0)q(Q^1, Q^0; P^1) \end{aligned}$$

But, in general,

$$(5.14)\quad \begin{aligned} p(P^1, P^0; Q^i)q(Q^1, Q^0; P^i) &\not\equiv e^i/e^0 \\ &\quad (i = 0, 1) \\ p(P^1, P^0, Q^a)q(Q^1, Q^0; P^\alpha) &\not\equiv e^1/e^0 \\ &\quad \text{for arbitrary } (Q^a, P^\alpha) \end{aligned}$$

From this review of the general case, it will be appreciated how much more simple is the homothetic case. If only it were as realistic[10] as it is elegant! This is said not

[10] Fortunately, in the case of production theory, because of the degree of interest that attaches to constant returns to scale, homotheticity is not always so unrealistic. See Ulmer, ch. 4, for an attempt to estimate limits even in the nonhomothetic case. More promising is the quadratic interpolation method of Abraham Wald that Ulmer discusses in his appendix A along with the ambitious attempt of Frisch (1932). Afriat favors the linear Engel-curve approximation: $e(P; Q) = \theta(P)\phi(Q) + \mu(P)$.

in regret, but as a reminder that more is asked of index numbers than can be delivered, as for example when union contracts and government sliding-scale agreements are all made to depend on the consumers price index appropriate at best to one real-income stratum.

VI. Divisia Indexes

Before leaving the general survey, a word about the little understood Divisia line integrals is in order. The typical argument for Divisia indexes is inadequate in its treatment of the invariance of the line integrals under change of path between endpoints. Too often it goes something like this Schumpeter version

$$(6.1) \qquad d\sum_1^n p_jq_j = \sum_1^n p_j dq_j + \sum_1^n q_j dp_j$$

or in vector notation

$$d(P\cdot Q) = P\cdot dQ + Q\cdot dP$$

$$P^1\cdot Q^1 - P^0\cdot Q^0 = \int_c P\cdot dQ + \int_c Q\cdot dP$$

$$= I_q + I_p$$

where c indicates the contour taken by $[p_j(\mathrm{t}), q_j(\mathrm{t})]$ between the endpoints, t^1 and t^0. However, for each different path going from the end points t^0 to t^1, the invariant left-hand difference gets broken down on the right-hand side into noninvariant partition between the I_q and I_p line integrals.

More sophisticatedly, beg the case of existence and suppose there somehow exist composite scalar measures of quantity and price, $\bar{q}$ and $\bar{p}$, such that

$$(6.2) \qquad \bar{q}\bar{p} = Q\cdot P, \quad \log \bar{q} + \log \bar{p} = \log \sum_1^n p_jq_j$$

$$\log Q^1\cdot P^1 - \log Q^0\cdot P^0$$

$$= \int_c \left[\sum_1^n p_jq_j\right]^{-1} \sum_1^n p_j dq_j$$

$$+ \int_c [\sum p_jq_j]^{-1} \sum q_j dp_j = J_q + J_p$$

$$\bar{q}^1 = \bar{q}^0 \exp J_q, \qquad \bar{p}^1 = \bar{p}^0 \exp J_p$$

But, again, this begs the case as to whether the invariant total $J_q + J_p$ is composed of separate line integrals that are invariant under arbitrary change of the $[Q(\mathrm{t}), P(\mathrm{t})]$ path, c, between the fixed endpoints. In general, from the theory of Pfaffians and inexact differential expressions, the integrability conditions needed to ensure invariant line integrals are *not* realized.

However, as seen in Samuelson (1965, equation (5)), for Y defined as P/e,

$$(6.3) \qquad Q = \mathit{grad}\ \log p(Y) = [\partial \log p(y_1, \ldots, y_n)/\partial y_i]$$

$$(6.3') \qquad Y = \mathit{grad}\ \log q(Q) = [\partial \log q(q_1, \ldots, q_n)/\partial q_i]$$

Hence, independently of path

$$(6.4) \qquad \log [q(Q^1)/q(Q^0)]$$

$$= \int_c \mathit{grad}\ \log q(Q)\cdot dQ = \int_c Y\cdot dQ$$

$$= \int_{\mathrm{t}^0}^{\mathrm{t}^1} \sum_1^n y_j(\mathrm{t}) q_j'(\mathrm{t}) d\mathrm{t}$$

$$= \int_{\mathrm{t}^0}^{\mathrm{t}^1} \sum_1^n [y_j(\mathrm{t}) q_j(\mathrm{t})] \cdot [d \log q_j(\mathrm{t})/d\mathrm{t}] d\mathrm{t} \equiv J_q$$

$$(6.4') \qquad \log [p(P^1)/p(P^0)]$$

$$= \log [p(Y^1)/p(Y^0)] + \log [e^1/e^0]$$

$$= \int_c \mathit{grad}\ \log p(Y)\cdot dY + \log [e^1/e^0]$$

$$= \int_c Q\cdot dY + \log [e^1/e^0]$$

$$= \int_{\mathrm{t}^0}^{\mathrm{t}^1} \sum_1^n [q_j(\mathrm{t}) y_j'(\mathrm{t})] d\mathrm{t} + \log [e^1/e^0]$$

$$= \int_{t^0}^{t^1} \sum_1^n [y_j(t)y_j(t)][d \log q_j(t)/dt]dt$$
$$+ \log [e^1/e^0]$$
$$= \int_{t^0}^{t^1} \sum_1^n [y_j(t)q_j(t)]$$
$$\cdot [d \log p_j(t)/dt]dt + 0$$
$$= J_p$$

This provides a proper proof that if and only if homotheticity prevails, we can define canonical index numbers by Divisia line integrals that will be invariant under arbitrary path change

$$(6.5)\quad q(Q^1)/q(Q^0) = \exp \int_{t^0}^{t^1} \left[\sum_1^n y_j(t)q_j(t) \right] \cdot [d \log q_j(t)/dt]dt$$

$$(6.5')\quad p(P^1)/p(P^0) = \exp \int_{t^0}^{t^1} \left[\sum_1^n y_j(t)q_j(t) \right] \cdot [d \log p_j(t)/dt]dt$$

If one's data happen to come along a continuous arc $[p_i(t), q_i(t)]$ and if homotheticity is a legitimate assumption (as it often may be in production theory), one may fit the surfaces $p(Y)$ and $q(Q)$ to the data by the above normalized Divisia-index expressions. This is not a better method of surface fitting than an arbitrary variety of other methods of the type to be discussed in later sections, but for the data given on a continuous arc, it is a convenient method and one that must be as good as any other.

When, as is invariably the case, our data are at discrete intervals, then even if they happen to come in a determinate time sequence, it is no longer necessarily the case that the best way of utilizing the data is to attempt to approximate the normalized Divisia integrals by one or another method of numerical integration (for example, Simpson's rule, Euler-Maclaurin formulas, etc.).[11]

Our investigation of the history of Divisia indexes is worth reporting. Schumpeter, chapter 8, gives the typical cavalier exposition that ignores the problem of lack of path invariance of the line integrals. But already in Divisia's original work in the 1920's, as in Divisia (1928) and in other references given in Dale Jorgenson and Zvi Griliches on the subject, the problem is ignored—as it is in the Frisch 1936 "Survey" and in the works by Dresch and most others in the G. C. Evans Berkeley camp (Shephard being a notable exception). The remarkable 1946 paper of Ville gives the first proof we could find that, only with homotheticity, do we get the vital path invariance. (Incidentally, our duality relations, (6.3) and (6.3′) above

[11] Actually, if the data are "close together," or the rate of change of events is very "slow," it will normally be better to use *all* the data points *simultaneously* in the fitting process as in equations (7.14) and (7.16) below. Thus, let $n=2$ and suppose you have N observations of vector pairs listed consecutively in time: $[y(t_1), q(t_1); y(t_2), q(t_2); \ldots ; y(t_N), q(t_N)]$. Then, without regard to t_i positioning, we can calculate N pairs of $[y_2/y_1, q_2/q_1]$ and from them fit the best functional relationship $y_2/y_1 = R(q_2/q_1)$ as in (7.13) below.

On an NSF grant, Meir Kohn of the M.I.T. Graduate School has kindly applied our ϵ-power series technique of Section VII to see whether the familiar device of "chaining" index numbers is an optimal mode of calculation. As suspected, calculating the Divisia integrals interval by interval by chaining does do better than merely using the pair of remote endpoints: as we show in (7.14) and later, the latter procedure gives an error-remainder term proportional to $T^3\epsilon^3$ whereas Kohn's Divisia-index chaining gives an error-remainder term proportional only to $T\epsilon^3$, where T is the number of equally spaced time intervals. Using *all* the $(T+1)$ vectors of data *simultaneously* should be able, certainly in the $n=2$ case of our discussion around equation (7.14), to annihilate error terms in $\epsilon^3, \epsilon^4, \ldots$, and in favorable cases even in ϵ^T. In unpublished work, Spencer Star and Robert E. Hall have shown that one gets a better discrete-data approximation to the Divisia line integrals when one assumes that, within each subinterval $\Delta t = t_{i+1} - t_i$, $y_j q_j$ is assumed constant (at the mean of its t_{i+1} and t_i values) than if one merely calculates $\int Y \cdot dQ$ for the interval by the approximation $Y(t_i) \cdot [Q(t_{i+1}) - Q(t_i)]$. Star and Hall give exact error calculations in terms of (not necessarily observable) covariance integrals; and, actually, in some realistic cases, the $T^3\epsilon^3$ coefficient may be small (even if not smaller than the $T\epsilon^3$ term).

are already in Ville, but the Houthakker Strong Axiom seems *not* to be *quite* there.) Wold, Richter, Gorman, Samuelson, Jorgenson and Griliches, Hulten, and other modern writers are aware of homotheticity's key role. In production theory, unlike consumer's demand analysis, this is not so bizarre a hypothesis: however, where the homothetic isoquants do not correspond to constant-returns output function, exp J_q gives at best *canonical* $q(Q^1)/q(Q^0)$ not $Output^1/Output^0$; moreover, one cannot then blithely identify p_i/e expressions with percentage marginal productivities $\partial \log Output/\partial q_i$.

VII. Approximations

In this paper we have thus far been concerned primarily with exact formulations and not with the problem of approximation. However, the two have become confused and so we do make some needed observations upon the problem of approximation and empirical estimation based upon incomplete information, particularly since the theory of revealed preference and maximization does provide some correct limits on exact functions.

Thus, from the definition of $e(P^1; Q^0)$ as minimized expenditure to achieve the $u(Q^0)$ level of satisfaction, we derive *alternative* one-sided bounds

$$(7.1)\quad p(P^1, P^0; Q^0) = \frac{e(P^1; Q^0)}{e(P^0; Q^0)} \leq \frac{P^1\cdot Q^0}{P^0\cdot Q^0} = \lambda_p$$

$$(7.1')\quad p(P^1, P^0; Q^1) = \frac{e(P^1; Q^1)}{e(P^0; Q^1)} \geq \frac{P^1\cdot Q^1}{P^0\cdot Q^1} = \pi_p$$

Here λ_p and π_p stand for the familiar Laspeyres and Paasche price indexes. Likewise we denote by λ_q and π_q the corresponding base-period weighted and given-period weighted quantity indexes with $\lambda_p\pi_q = e^1/e^0 = \lambda_q\pi_p$. Then, by symmetry, similar alternative one-sided limits hold for quantity indexes, namely

$$(7.2)\quad q(Q^1, Q^0; P^0) \leq \frac{P^0\cdot Q^1}{P^0\cdot Q^0} = \lambda_q$$

$$(7.2')\quad q(Q^1, Q^0; P^1) \leq \frac{P^1\cdot Q^1}{P^1\cdot Q^0} = \pi_q$$

As became clarified in the 1930's it is in general invalid to combine the above alternative single-limit expressions into the frequently met assertion

$$(7.3)\quad \pi_p \leq p(P^1, P^0; Q^i) \leq \lambda_p$$

Only in the special homothetic case in which

$$p(P^1, P^0; Q^1) \equiv p(P^1, P^0; Q^0) = p(P^1)/p(P^0)$$

can we validly[12] derive the following double limit

$$(7.4)\quad \pi_p \leq p(P^1, P^0) = p(P^1)/p(P^0) \leq \lambda_p$$

$$(7.4')\quad \pi_q \leq q(Q^1, Q^0) = q(Q^1)/q(Q^0) \leq \lambda_q$$

It is true that for the general case one can almost trivially assert a double limit of the following form

$$(7.5)\quad \text{Min}\,[\ldots, \overset{1}{p_j}/\overset{0}{p_j}, \ldots] \leq p(\overset{1}{P}, \overset{0}{P}; \overset{0}{Q}) \leq \lambda_p \leq \text{Max}\,[\ldots, \overset{1}{p_j}, \overset{0}{p_j}, \ldots]$$

and likewise for $p(P^1, P^0; Q^i)$ with λ_p

[12] Franklin Fisher and Karl Shell for the most part invoke the nonvalid double-limit (7.4), acknowledging in their fn. to p. 6, fn. 7 on p. 38, that it is not generally valid. But the reader must be warned against their reason for why the double-bound is not generally valid. Instead of attributing its invalidity to nonhomotheticity, they say: "In fact this [double-bound] proposition is not true [even with unchanged tastes] if price and income changes are large." It is incorrect that *largeness* of (P^1-P^0, Q^1-Q^0) has anything to do with the failure of the double-bound. Even if these changes were indefinitely small, it would not be true that $\lambda_q \leq q(Q^1, Q^0; p_i) \leq \Pi_q$, as the failure of (iv*) in our earlier discussion of (5.7) reveals and as (7.25) will discuss further for the example $Q^1 = Q^0(1+\epsilon)$, $|\epsilon|$ arbitrarily small.

omitted. These follow from the easily demonstrated fact that a price index number is a *generalized internal mean* of its price ratios $[\ldots, p_j^1/p_j^0, \ldots]$. But these double limits are often so wide as to be practically not worth very much. Thus, in the homothetic case, (7.4) generally gives a sharper bound on the left than (7.5). And, as was pointed out in connection with (5.6) and (i*) interchanging all q_j for p_j in (7.5) will *not* in the nonhomothetic case give a valid (7.5′) that provides similar bounds on $q(Q^1, Q^0; P^\alpha)$.

Another odd confusion crops up repeatedly in the literature. *If* we know that $u(Q^1)=u(Q^0)$, then it is valid to write the two-sided bounds

$$\text{(7.6)}\qquad \pi_p \le p(P^1, P^0; Q^\alpha) \le \lambda_p, \qquad u(Q^1) = u(Q^0) = u(Q^\alpha)$$
$$\pi_q \le q(Q^1, Q^0; P^\alpha) = 1 \le \lambda_q \quad \text{for all } P^\alpha$$

But since we know $q=1$ by hypothesis, no bounds on q are needed! Nor do the bounds on p add any information since, in this case, we *already* know the *exact* value of p as e^1/e^0!

Let us return to the homothetic case in which (7.4) and (7.4′) are valid double-bounds. Is it not tempting to take some symmetric mean of the upper and lower bounds of Paasche and Laspeyres in the hope of getting a "more accurate approximation" and avoiding "upward or downward bias"? Fisher's Ideal index already mentioned in footnote 9, which in our notation is $(\lambda_p\pi_p)^{1/2}$, is precisely such an attempt. And, even though it fails the general circular test, we would welcome it as an approximation if in some sense it tended to produce "more accurate, less biased" results.

At long last, after half a century, we can a little bit vindicate Fisher in his choice of the "Ideal" index number, $(\lambda_p\pi_p)^{1/2}$ or $(\lambda_q\pi_q)^{1/2}$, not as an exact formula for $p(P^1)/p(P^0)$ or $q(Q^1)/q(Q^0)$—for there can be no one formula to cover two different sets of tastes—but as one "locally sufficient" function of *homothetic* (P, Q) observations that will approximate to the true magnitude up to terms of the *second* degree in the deviations between the two situations. To interpret what this means, we now proceed to develop a new power-series technique that enables one to make rigorous sense about degrees of goodness of approximations.

Specifically, for $Y=P/e$ and the two situations close together in *ratios*, so that $Q^1-\alpha Q^0$ and $Y^1-\alpha^{-1}Y^0$ are "small numbers," write

$$\text{(7.7)}\qquad \begin{aligned} \alpha Y^1 &= Y^0 + \epsilon \dot{Y}^0 + \tfrac{1}{2}\epsilon^2 \ddot{Y}^0 + \ldots \\ \alpha^{-1}Q^1 &= Q^0 + \epsilon\dot{Q}^0 + \tfrac{1}{2}\epsilon^2\ddot{Q}^0 + \ldots \\ Y^1\cdot Q^1 &= 1 + 0 = Y^0\cdot Q^0 \\ &= 1 + \epsilon(Y^0\cdot\dot{Q}^0 + \dot{Y}^0\cdot Q^0) \\ &\quad + \tfrac{1}{2}\epsilon^2(Y^0\cdot\ddot{Q}^0 + 2\dot{Y}^0\cdot\dot{Q}^0 \\ &\quad + \ddot{Y}^0\cdot Q^0) + \ldots \\ Y^0\cdot\dot{Q}^0 &= -\dot{Y}^0\cdot Q^0, \\ -\ddot{Y}^0\cdot Q^0 &= Y^0\cdot\ddot{Q}^0 + 2\dot{Y}^0\cdot\dot{Q}^0, \text{ etc.} \end{aligned}$$

From (6.3′), we know that

$$\text{(7.8)}\qquad \partial q(Q)/\partial q_i = q(Q)y_i \quad (i = 1, \ldots, n)$$

Substituting (7.7) and (7.8) into $q(Q^1)$ gives

$$\text{(7.9)}\qquad \begin{aligned} &\alpha^{-1}q(Q^1)/q(Q^0) \\ &= 1+\epsilon Y^0\cdot\dot{Q}^0+\tfrac{1}{2}\epsilon^2[Y^0\cdot\ddot{Q}^0+2Y^0\cdot Q^0 \\ &\quad +(\dot{Y}^0\cdot Q^0)^2]+\ldots \end{aligned}$$

By (6.3′) and the definitions of λ_q and π_q

$$\text{(7.10)}\qquad \begin{aligned} \alpha^{-1}\lambda_q &= \alpha^{-1}Y^0\cdot Q^1 = 1+\epsilon Y^0\cdot\dot{Q}^0 \\ &\quad +\tfrac{1}{2}\epsilon^2 Y^0\cdot\ddot{Q}^0+\ldots \\ \alpha^{-1}\pi_q &= \alpha^{-1}(Y^1\cdot Q^0)^{-1} = \{1+\epsilon\dot{Y}^0\cdot Q^0 \\ &\quad +\tfrac{1}{2}\epsilon^2\ddot{Y}^0\cdot Q^0+\ldots\}^{-1} \\ &= 1-\epsilon\dot{Y}^0\cdot Q^0+\tfrac{1}{2}\epsilon^2[2(Y^0\cdot\dot{Q}^0)^2 \\ &\quad -\ddot{Y}^0\cdot Q^0]+\ldots \\ &= 1+\epsilon Y^0\cdot\dot{Q}^0+\tfrac{1}{2}\epsilon^2[2(Y^0\cdot Q^0)^2 \\ &\quad +2\dot{Y}^0\cdot\dot{Q}^0+Y^0\cdot\ddot{Q}^0]+\ldots \end{aligned}$$

$$(7.10') \quad \alpha^{-1}(\pi_q\lambda_q)^{1/2} = 1+\epsilon Y^0\cdot\dot{Q}^0+\tfrac{1}{2}\epsilon^2[Y^0\cdot\ddot{Q}^0 + \dot{Y}^0\cdot\dot{Q}^0+(\dot{Y}^0\cdot Q^0)^2]+\ldots$$

But this last expression, whose first three coefficients are simple linear unweighted means of the λ_q and π_q expressions' coefficients, is seen to be identical up to ϵ^2 terms with the true index of (7.9). Hence, we have established the following:

ACCURACY THEOREM: *In the homothetic case, any symmetric mean of the Laspeyres and Paasche index numbers (including the Ideal index's geometric mean) will approximate the true index number up to the third order in accuracy.*

$$(7.11) \quad q(Q^1)/q(Q^0) - (\lambda_q\pi_q)^{1/2} = 0 + \epsilon 0 + 0\epsilon^2 + \epsilon^3 r(\epsilon)$$

where the remainder $r(\epsilon)$ is finite at $\epsilon=0$.

The truth of this finding, that the Ideal index gives a second-order or osculating approximation to the true homothetic index, could have been vaguely suspected from the finding in Samuelson (1953a, p. 8, n. 1) that the symmetric mean of overcompensated and undercompensated demand functions provides a high-order, osculating approximation to the Slutsky-Hicks just-compensated demand along the indifference contours.

The Ideal index is of course not alone in this accuracy. *Any* symmetric internal mean of the locally sufficient Laspeyres and Paasche indexes will provide as high accuracy. For example, $p(P^1)/p(P^0)$ is well-approximated by any generalized symmetric mean function $m(\lambda_p, \pi_p)$, where

$$(7.12) \quad \text{Min }(x, y) \leq m(x, y) \equiv m(y, x) \leq \text{Max }(x, y)$$

One would have to go to ϵ^3 terms to decide whether one symmetric mean can be said to be better than another, as for example whether $\frac{1}{2}(\lambda_q+\pi_q)$ gives a better approximation than $(\lambda_q\pi_q)^{1/2}$ to a particular given $q(Q)$.

Of course, all this applies only to the homothetic case, as we shall see.

The fact that the Ideal index, for Q and P/e treated as independently varying vectors, flunks the circular test does not vitiate it as a local approximation. Approximations often violate transitivity. For example, 1.01 and .99 are each within 1 percent approximations to 1.0, but *that* does not make them have this property with respect to each other!

The Homothetic Case with Only Two Goods

Given *only* two observations $[P^\beta, Q^\beta]$, $(\beta=0, 1)$, no more information can be squeezed out about the unknown $q(Q)$ or $p(P)$ functions than from these $m(\lambda, \pi)$ functions. But if we have more than two observations, $(\beta=0, 1, 2, \ldots)$, say k in all, we can of course do better. It appears that the independent inner-products $P^\beta\cdot Q^\alpha$ are no longer "locally sufficient" information parameters for highest-order accuracy. Let us sketch briefly for the $n=2$ case how k observations might give us approximations to $q(Q)$ accurate up to ϵ^k terms. Write

$$(7.13) \quad q(q_1, q_2)=q_1q(1, q_2/q_1)=q_1g(q_2/q_1)$$
$$p_1/p_2=\frac{g(q_2/q_1)-(q_2/q_1)g'(q_2/q_1)}{g'(q_2/q_1)} = \exp-h[\log(q_2/q_1)]$$

Now from k nearby paired values of $[\log q_2^k/q_1^k, \log (p_2/p_1)^k]$, fit a polynomial approximation to $h[\ ]$, namely

$$(7.14) \quad h[z] = h + h_1z + \ldots + h_{k-1}z_{k-1}$$

Knowing $h[z]$, we use the differential equation for $g(\ )$ in (7.13) to get its corresponding form, and then we have $q(q_1, q_2) = q_1g(q_2/q_1)$ and can calculate our index numbers of high-order accuracy, by $q(Q^\beta)/q(Q^0)$.

As mentioned in footnote 11 on Divisia index chaining, even when our data come in consecutive time or space sequence, one does not get an optimal estimate of an existent $q(Q)$ function by using binary chaining, in which one estimates $q(Q^T)/q(Q^0)$ by chained separate products $[q(Q^1)/q(Q^0)][q(Q^2)/q(Q^1)] \ldots [q(Q^T)/q(Q^{T-1})]$. Instead, when the data are sufficiently close and each $Q^t = Q^0 + \epsilon\dot{Q}^0 + \ldots$, one gets a lower coefficient of the ϵ^3 and higher-order error terms by using all data simultaneously to make closest estimates of the coefficients in

$$(7.15) \quad q[\epsilon] = q(Q[\epsilon]) = q[0] + q'[0]\epsilon + q''[0]\epsilon^2/2 + \ldots$$

The Homothetic Case with More Goods

For $n > 2$, the problem is of course more complicated, even in this homothetic case. Only a sketch can be given here. For the homothetic case we are fortunate in that the Antonelli-Slutsky-Hotelling-Hicks integrability conditions take a very simple form and enable us to reduce the problem to that of determining one function of n variables: $\phi(Q) = log\ q(Q)$ being the most convenient one. It was seen in (6.3′) and (7.8) that $Y = grad\ \phi(Q) = (\partial\phi(Q)/\partial q_j)$. Because $q(Q)$ is first-degree homogeneous, we can reduce the problem to that of determining a surface in the $n-1$ variables $(q_j/q_1) = (z_2, \ldots, z_n) = Z$, namely $\phi(1, Z) = \psi(Z)$ whose gradient equals the modified prices $W = (w_2, \ldots, w_n) = (q_1y_2, \ldots, q_1y_n)$.

Now suppose we are given for $t = 0, 1, \ldots, T$ data situations the observed vector pairs $[Z^t, W^t] = [z_2^t, \ldots, z_n^t; w_2^t, \ldots, w_n^t]$. Then we can fit a polynomial surface to $\psi(Z)$ of the form

$$(7.16) \quad h(z_2, \ldots, z_n) = a_0 + \sum_{j=1}^{n} a_{1j}(z_j - \overset{0}{z_j}) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{2,ij}(z_i - \overset{0}{z_i})(z_j - \overset{0}{z_j}) + R_3$$

$$(7.17) \quad grad\ h(\overset{0}{Z}) = \overset{0}{W} = [\overset{0}{w_j}] = [a_{1j}]$$

Thus, to determine only $n-1$ coefficients a_{1j} to $h(Z)$, we need only one (Z^0, W^0) observation. But for a quadratic approximation to $h(Z)$, we must determine the $n-1$ coefficients a_{1j} plus $n(n-1)/2$ independent $a_{2,ij}$ coefficients: this requires $(n-1)\cdot[1+\frac{1}{2}n]$ equations in all; but since, for each t in (7.17), we get $n-1$ equations, we need at least $T = 1 + \frac{1}{2}n$ situations to be able to derive a quadratic approximation. Thus, for $n-1 = 1$, and T situations we can, as in (7.14) determine $\psi(Z)$ as a T degree polynomial. For $n-1 = 2$, we would need to determine 5 coefficients for a quadratic approximation; hence $T = 3$ known values for $[(q_2/q_1)^t, (q_2/q_1)^t; (q_1y_2)^t, (q_1y_3)^t]$ would give us 6 equations, which is 1 more than the minimum number needed. For general $n-1$, we need approximately $n/2$ equations to get a quadratic approximation. Note, however, that for data close together so that the (7.12) power series applies, we can with a large enough number of situations T, in principle, determine the $q(Q)$ surface to as great accuracy as we desire. Of course, the T situations must not be collinear, as when all (q_i/q_1) proportions never change at all. And, as is known from the theory of interpolation of noisy data, one might want to use redundant observations to filter out error noise and make best estimates of true message, utilizing least squares and other techniques for providing data. If the $q(Q)$ and $\psi[Z]$ functions are not indefinitely differentiable, or if their higher derivatives oscillate wildly, one will prefer low-degree polynomial fitting to high.

Before leaving this little explored domain, we should mention that when $q(Q)$ is actually changing through time—as in our Section VIII dealing with changing tastes—chaining makes some sense, even though it may be hard to state what then makes good sense.

The Nonhomothetic Two-Good Case

The nonhomothetic case also presents, even for $n=2$, more intrinsic difficulties. With only 2 observations, we do not perceive that there is *any* way of getting $q(Q^1, Q^0; P^\alpha)$ to an accuracy of ϵ^2. However, given 3 nearby and noncollinear observations on (q_1, q_2, y_1, y_2), we should be able to do so. Thus, let

$$(7.18)\qquad Q^t = Q^0 + \epsilon A^t, \qquad (t = 1, 2)$$
$$det\,[A^1, A^2] \neq 0$$

and suppose we also have observations on (Y^2, Y^1, Y^0). Then in

$$(7.19)\qquad y_2/y_1 = [\partial u(Q)/\partial q_2]/[\partial u(Q)/\partial q_1] = r(q_1, q_2)$$

we can use our three paired observations $[q_1^t, q_2^t, (y_2/y_1)^t]$, $(t=0, 1, 2)$, to provide some approximation to the marginal rate of substitution function, $r(\ ,\ )$, that is linear in terms of some functions of the variables.

Often, it is alternatively assumed, as for example in Wald, that we have knowledge of two or more Engel's paths from budget studies in different price situations. Then we know two empirical functions, $f[q_2; (p_2/p_1)^t]$ in the following

$$(7.20)\qquad r(q_1, q_2) = (p_2/p_1)^t \qquad (t = 0, 1)$$
$$q_1 = f[q_2; (p_2/p_1)^t]$$

and can use linear interpolation, either on the q's, their logarithms, or any other stretchings, to get the $f[\ ;\ ]$ and $r(\ ,\ )$ functions up to accuracy of terms $\{(p_2/p_1)^1 - (p_2/p_1)^0\}^2$.

Given the $r(\ ,\ )$ function or a close approximation to it, we then, as in Samuelson (1948, p. 245), solve the differential equation

$$(7.21)\qquad -dq_2/dq_1 = r(q_1, q_2)$$

to get

$$(7.22)\qquad q_2 = u(q_1; \overset{\#}{q_2}), \qquad \overset{\#}{q_2} \equiv v(\overset{0}{q_1}; \overset{\#}{q_2})$$
$$\overset{\#}{q_2} = v^{-1}(q_1; q_2) = f\{u(Q)\}$$

Having derived one indicator of utility, our index number problems are solved.

The Nonhomothetic Many-Good Case

Here we shall be brief and avoid duplicating the exposition of Wald, and shall not be assuming knowledge of any full Engel's paths. Write

$$(7.23)\qquad u(Q) = (Q - Q^0)\begin{bmatrix}1\\ b_2\\ \cdot\\ \cdot\\ b_n\end{bmatrix} + \frac{1}{2}(Q - Q^0)\cdot[a_{ij}](Q - Q^0)^T + R_3$$

Although we shall here neglect R_3 remainders and speak only of quadratic approximations, our method generalizes to cubic and higher-degree approximations. Note that we have chosen to employ the convention of setting $u(Q^0)=0=1-\partial u(Q^0)/\partial q_1$. By the familiar tangency of price ratios and indifference contours, we know

$$(7.24)\qquad \frac{\partial u(Q^t)/\partial q_i}{\partial u(Q^t)/\partial q_1} = (p_i/p_1)^t, \qquad (t = 0, 1, \ldots, T;\ i = 2, \ldots, n)$$
$$0 = \partial u(Q^t)/\partial q_i - (p_i/p_1)^t \partial u(Q^t)/\partial q_1$$
$$0 = b_i - (p_i/p_1)^0$$
$$0 = \left[(p_i/p_1)^0 - \sum_1^n a_{ij}q_j^t\right] - (p_i/p_1)^t\left(1 - \sum_1^n a_{1j}q_j^t\right)$$

These last are $(n-1)T$ relations linear in the unknown a_{ij} coefficients, of which, because of the symmetry of the a_{ij} coefficients, there are only $n(n+1)/2$ unknown a_{ij} coefficients. This means that for $T>\frac{1}{2}(n-1)+1.5$ and no degeneracy or

$$
\begin{aligned}
(7.25)\qquad u(Q) &= q_2(1+q_1), \qquad Q^0 = (1, 1), \qquad Q^1 = (1+\epsilon, 1+\epsilon)\\
(p_2/p_1)^0 &= (1+1)/1 = 2, \qquad (p_2/p_1)^1 = (2+\epsilon)/(1+\epsilon)\\
e(P^\alpha; Q^1) &= \operatorname*{Min}_{Q}\,\{q_1 + (p_2/p_1)^\alpha q_2\} \text{ such that } q_2(1+q_1) = (1+\epsilon)(2+\epsilon)\\
e(p^\alpha; Q^1) &= \operatorname*{Min}_{q_1}\left\{q_1 + \frac{(p_2/p_1)(1+\epsilon)(2+\epsilon)}{1+q_1}\right\}\\
&= 2[(p_2/p_1)^\alpha(1+\epsilon)(2+\epsilon)]^{1/2} - 1\\
&= 2[2(p_2/p_1)^\alpha]^{1/2}\left[1+\frac{3}{2}\epsilon+\frac{\epsilon^2}{2}\right]^{1/2} - 1\\
&= 2[2(p_2/p_1)^\alpha]^{1/2}\left[1+\frac{3}{4}\epsilon-\frac{1}{32}\epsilon^2+\dots\right] - 1\\
q(Q^1, Q^0; P^\alpha) &= \frac{2[2(p_2/p_1)^\alpha]^{1/2}\left[1+\frac{3}{4}\epsilon-\frac{1}{32}\epsilon^2+\dots\right] - 1}{2[2(p_2/p_1)^\alpha]^{1/2} - 1}
\end{aligned}
$$

collinearity, we can in principle calculate all our desired functions: $u(Q)$, $q(Q, Q^0; P^\alpha)$, $p(P, P^0; Q^\alpha)$. If the T observations obey the power-series form of (7.12) generalized to nonhomotheticity, then for $\epsilon \to 0$ but $\epsilon \neq 0$, we can calculate the $q(\ ,\ ;\ )$ and $p(\ ,\ ;\ )$ index numbers that will be accurate up to powers of ϵ as high as we like—if only we are given T sufficiently large.

Nonoptimality of Ideal Index for Nonhomothetic Case

We have already noted that, in the nonhomothetic case, the quantity index can fail Fisher's first test of being a generalized mean of the ratios (q_j^1/q_j^0), and can differ from them even when they all have a common value. Since $(\lambda_q\pi_q)^{1/2}$ will equal that common ratio, it is evident that the Ideal index cannot give high-powered approximation to the true index in the general, nonhomothetic case. A simple example will illustrate the degree of this failure as shown in (7.25). Specifically, for $(p_2/p_1)^\alpha = (p_2/p_1)^0$,

$$
\begin{aligned}
(7.26)\quad q(Q^1, Q^0; P^0) &= \frac{2[2]\left[1+\frac{3}{4}\epsilon-\frac{1}{32}\epsilon^2+\dots\right]-1}{2[2]-1}\\
&= 1+\epsilon-\frac{1}{24}\epsilon^2+\dots\\
&< \lambda_q = 1+\epsilon = \pi_q = (\lambda_q\pi_q)^{1/2}
\end{aligned}
$$

for $\epsilon \neq 0$ but sufficiently small.

It is thus seen that there is *not* agreement to the second-order of ϵ^2 terms. This example also warns against the common fallacy mentioned in footnote 12: even if (P^1, P^0, P^α) and (Q^1, Q^0) are "sufficiently close together," it is not true that the Laspeyres and Paasche indexes provide two-sided bounds for the true index. In this example, the true index lies *outside* the $[\lambda_q, \pi_q]$ interval!

VIII. Taste Changes

As in Samuelson[13] and Fisher-Shell, we

[13] Samuelson in present notation, says: "We can validly state: $p>1$ [*i.e.*, $\pi_q>1$] implies that the first batch of goods is higher than the batch II on the indifference curves that prevailed in period I [*i.e.*, the

show what can be said about index numbers when tastes change, being generated in situations 0 and 1 not by uniform $u(Q)$ but respectively by $u^0(Q)$ and $u^1(Q)$. We now generalize the $e(\, , \, ; \,)$, $p(\, , \, ; \,)$, and $q(\, , \, ; \,)$ functions for $(i=0, 1)$ to

$$(8.1) \qquad e^i(P; Q) = \min_X P \cdot X$$

such that $u^i(X) = u^i(Q), \qquad (i=0, 1)$

$$p^i(P^1, P^0; Q^a) = e^i(P^1; Q^a)/e^i(P^0; Q^a)$$
$$q^i(Q^1, Q^0; P^\alpha) = e^i(P^\alpha; Q^1)/e^i(P^\alpha; Q^0)$$
$$e^i(P^j; Q^j) = P^j \cdot Q^{ji}$$
$$\neq e^j = P^j \cdot Q^j, \ (j \neq i)$$

Note that $p^i(P^1, P^0; Q^i)$ never involves Q^j at all, and neither do ratios involving $P^i \cdot Q^i$ and $P^i \cdot Q^j$. So the valid one-sided price bounds of (7.1) remain valid.

$$(8.2) \qquad p^1(P^1, P^0; Q^1) \geq \pi_p$$
$$(8.2') \qquad p^0(P^1, P^0; Q^0) \leq \lambda_p$$

Likewise, since neither $q^i(Q^1, Q^0; P^i)$ nor ratios involving such terms as $P^i \cdot Q^i$ and $P^i \cdot Q^j$ ever involve P^j at all (or the new and unobserved P^{ji} that would have to prevail *if* Q^j were to be the best buy at $u^i(Q)$ tastes), the one-sided quantity bounds of (7.2) remain valid:

$$(8.3) \qquad q^1(Q^1, Q^0; P^1) \geq \pi_q$$
$$(8.3') \qquad q^0(Q^1, Q^0; P^0) \leq \lambda_q$$

However, even in the homothetic case where $e^i(P; Q) = p^i(P)q^i(Q)$, the two-sided limits of (7.1) can never validly apply, being replaced only by

$$(8.4) \qquad \lambda_p \leq p^0(P^1)/p^0(P^0)$$
$$\leq \frac{P^1 \cdot Q^{10}}{P^0 \cdot Q^0} \gtrless \frac{P^1 Q^1}{P^0 Q^0} = \pi_p$$
$$\lambda_p = \frac{P^1 \cdot Q^0}{P^0 \cdot Q^0} \gtrless \frac{P^0 \cdot Q^1}{P^0 \cdot Q^{01}}$$
$$\leq p^1(P^1)/p^1(P^0) \leq \pi_p$$

Here Q^{01} are what would have been bought at P^0 if tastes of 1 had prevailed, etc.

Even the weak factor-reversal fails to be satisfied under changing tastes, since

$$(8.5) \qquad p^i(P^1, P^0; Q^1)q^i(Q^1, Q^0; P^0) \gtrless e^1/e^0 \qquad (i = 0, 1)$$
$$p^i(P^1, P^0; Q^0)q^i(Q^1, Q^0; P^1) \gtrless e^1/e^0$$

This failure greatly reduces the significance of the valid (8.2) price bounds since they can never be of any use in providing deflators of money expenditures that will produce estimates of real quantity magnitudes.[14] So it is fortunate that we have the valid direct quantity bounds of (8.3).

One use for (8.2) is in estimating cost of living allowances or adjustments for, say, foreign employees who face P^1 and need to be given enough (unknown) income, e_x^1, to be able to enjoy, at 0 situation tastes, a level of living just as good as Q^0. Then (8.3) provides the valid bound on e_x^1,

$$(8.6) \qquad e_x \stackrel{\text{def.}}{=} e^0 p^0(P^1, P^0; Q^0)$$
$$\leq e^0 \pi_p = P^0 \cdot Q^0 \frac{P^1 \cdot Q^0}{P^0 \cdot Q^0}$$
$$\leq P^1 \cdot Q^0$$

This solution is so simple as to be almost anticlimactic. It merely restates the commonsense observation that, if you give someone abroad enough money to buy

period of P^0, Q^0]" (1950, p. 24). Since $q(Q^1, Q^0; P)$ has just been shown to be approximated by λ_q up to terms of linear order in ϵ, if we are interested in the tastes of the period of (P^1, Q^1), we are well advised to follow the (8.3) suggestion above and give primacy to the Paasche π_q index. If, as is unlikely, we knew that tastes shift slowly—like ϵ^3 rather than ϵ—we would do better in the homothetic case to use a symmetric mean of λ_q and π_q, achieving ϵ^2 rather than ϵ accuracy.

[14] Fisher-Shell, at the end of their preface's first page, seem to suggest that the price index bounds they reach, such as (8.2)'s Paasche, can be used as deflators—which may be in some contradiction to (8.5), unless the deflation process is restricted to be of (8.6) type.

exactly the Q^0 goods he consumes at home, then he'll be at least as well off as he was at home, a so-called "over-compensated" change.

For two-country real-income comparisons, both tastes provide relevant $q^i(\, , \, ; \,)$ indexes. Often estimates of quantity changes are desired over more than two periods (so that one can spot acceleration, takeoffs, revolutions, or mundane cycles). If tastes are changing cumulatively (with or without acceleration, takeoffs, or cycles), it would not seem sensible to give primacy to the last-available year's tastes. This would involve recomputing every year against the new year; and such a procedure would ignore the useful information contained in the presumption that, between each two close-by periods, tastes have presumably not changed much. So a chaining procedure seems sensible. Thus, disregarding the primacy established here of relations like (8.2), one might well compute between successive periods the Ideal quantity indexes, chaining them together by appropriation multiplications; or one could use Divisia-like indexes if nonhomotheticity is not too great;[15] or one could use one or another of the improved approximations developed for unchanging tastes in our earlier sections.

Suppose the purpose is *to calculate average price changes* so that money wage rates and other contracts can be periodically "escalated." The vague and implicit purpose underlying such contracts is presumably that the worker is to be "as well off in Period 1 (or the period just following it) as he was in the base Period 0." This sounds almost as if Period 0's tastes are to provide the appropriate criterion, so that (8.3′)'s Laspeyres bound is appropriate rather than (8.3)'s Paasche bound.

Probably, though, one should not try to read anything so definite into people's vague notions of equity. When a General Motors cost-of-living contract continues in effect for year after year, it seems doubtful that primacy should be given to the bygone tastes of its first year of life. More likely, we shall want to concentrate on the tastes of its year or 3-year renewal. But, in so concentrating, there seems no reason why either the end-of-the-year or beginning-of-the-year tastes should be particularly favored. This again suggests chaining, perhaps in the even-handed form of multiplying together successively computed Ideal price indexes of adjacent periods.

In concluding this section, we note that quality changes in the vectors of goods Q^0 and Q^1 provide some further complications to make the allowances for taste changes even more difficult. As an example, suppose the later period adds a good q_{n+1} that was not available at any price in the earlier period. We must set its price then at infinity: $p^0_{n+1} = \infty$.

This makes our only valid bound on the price index calculated under the later period's taste, (8.2) involving the Paasche index, so wide as to be useless:

$$(8.7) \quad p^1(P^1, P^0; Q^1) \geq P^1 \cdot Q^1 / P^0 \cdot Q^0 = 0$$

since infinite p^0_j makes the last denominator infinite. Fortunately, our valid Paasche-quantity bound still does apply: it makes sense to ask, "How much could my later-period's income drop to leave me as well off as I would be with my present tastes if I had to consume today the Q^0 I used to consume?" Now we apply (8.2′) and give the answer: "If my income fell so that I could still buy Q^0 today, I'd assuredly still be as well off as then."

If, as following 400 A.D., goods both get lost and discovered, then no generally valid and useful bounds apply to prices or

[15] If one knows something about Engel-law deviations from homotheticity and something about Gerschenkron-Clark-Kuznets laws of changing comparative advantages with development, one should be able to improve upon the Ideal or Divisia indexes with their homotheticity dependence for high-order accuracy. This analysis is developed in Samuelson (1974b).

to quantities under any pattern of tastes (uniform or changing). That's the way things are.

IX. Index Numbers of Production Possibilities

Writers such as Hicks, Kuznets, and Little are cited in Samuelson (1950) where the focus is shifted away from cost-of-living consumption index numbers based on indifference contours to the attempt to infer uniform improvements in production-possibility frontiers. See also Abram Bergson, Richard Moorsteen, and Fisher-Shell, Essay II, for other more recent discussions.

Thus, let $X=(x_1, \ldots, x_n)$ now stand for amounts of goods produced. Let the frontier of what is producible be, for fixed totals of factor inputs and technology in the background, written as the implicit transformation function

$$(9.1) \quad 0 = f(X; x) = f(x_1, \ldots, x_n; x), \qquad \partial f/\partial x_j > 0 \quad (j = 1, \ldots, n)$$

where x is a scalar parameter of aggregate size that plays a role much like that played in demand theory by q in the $q(Q)$ function. Assume first that $\partial f/\partial x$ is known to be negative; an increase in x can then be seen to shift the p-p frontier outward, making more of all goods producible.

If $f(\ )$ is convex (from below) in all its X arguments (so that its contours in the $(x_1, x_2, \ldots)$ space for fixed x are concave!) and if $\partial f/\partial x$ is one-signed, an increase in x must increase the maximum national income attainable for any fixed set of positive prices. That is,

$$(9.2) \qquad y(Y; x) = \underset{X}{\text{Max}}\ Y \cdot X$$

subject to $f(X; x) = 0$

This is exactly parallel to the fact in consumption theory that minimized $P \cdot Q$ subject to fixed level of u is a monotone function of the utility indicator u. Therefore, to compare the production possibilities of two situations, for which we have observations (X^1, Y^1) and (X^0, Y^0), we should be able to define index numbers involving $Y \cdot X$ ratios that run exactly parallel to the consumption index numbers. Specifically, define for the index of quantity change

$$(9.3) \quad x(X^1, X^0; Y^\alpha) = y(Y^\alpha; X^1)/(Y^\alpha; X^0)$$

Just as the cost-of-living case is enormously simpler in the homothetic case, so can we get richer analytic results in the special "neutral-technical-change or homothetic case" where (8.1) can be specialized to the form

$$(9.4) \qquad x = x(x_1, \ldots, x_n) = x(X) = \lambda^{-1}x(\lambda X)$$

but where, because of the law of diminishing returns, $x(X)$ is convex, homogeneous first-degree and not concave (as in the consumer's diminishing marginal rate of substitution case). Reminder: The convexity of the $x(X)$ function should not be confused with the concavity of its contours in the $(x_1, x_2, \ldots)$ space.

It will come as no surprise to anyone who has followed our analysis thus far that running parallel to the homothetic consumption theory of price and quantity index numbers and their duality relations is a homothetic production-theory version of price and quantity index numbers. (The only difference is that convexity leads to maximization where concavity led to minimization and all bounds are reversed as in Samuelson (1950).)

Tersely, we define the dual to $x(X)$ as

$$(9.5) \qquad y(Y) = \underset{X}{\text{Max}}\ Y \cdot X/x(X)$$

$$y(Y)x(X) \geq Y \cdot X$$

with the equality sign holding if, and only if, the (X, Y) satisfy optimizing demand (or, better, "supply") pairings. As quantity and price index numbers, we define

$$x(X^1, X^0) = x(X^1)/x(X^0) \equiv x(X^1, X^0; Y^\alpha)$$
$$y(Y^1, Y^0) = y(Y^1)/y(Y^0) \equiv y(Y^1, Y^0; P^\alpha)$$

for *any* Y^α and x.

These satisfy all Fisher's suitably-generalized tests, having the properties

(9.6)

(i) $y(\lambda Y^0, Y^0) \equiv \lambda$

(ii) $y(Y^1, Y^0)y(Y^0, Y^1) = 1$

(iii) $y(Y^2, Y^1)y(Y^1, Y^0)y(Y^0, Y^2) = 1$

(iv) $y\dagger(Y\dagger^1/Y\dagger^0) \equiv y(Y^1, Y^0)$ for $y_j\dagger = y_j d_j^{-1}, \quad d_j > 0$

Similarly, $x(X^1, X^0)$ satisfies the time reversal, circular reversal, and dimensional invariance tests and provides a generalized internal mean of the x_j^1/x_j^0 ratios; hence the (9.6) tests are passed by quantity as well as price indexes.

The (weak) factor-reversal test is also, by duality of (8.4), satisfied.

(9.7) $y(Y^1, Y^0)x(X^1, X^0) = Y^1 \cdot X^1/Y^0 \cdot X^0$

Except that convexity on the supply functions makes us reverse the double Laspeyres-Paasche inequalities, we have just as in (7.4) and (7.4′)

$$\lambda_y = Y^1 \cdot X^0/Y^0 \cdot X^0 \leq y(Y^1, Y^0) \leq Y^1 \cdot X^1/Y^0 \cdot X^1 = \pi_y$$
$$\lambda_q = Y^0 \cdot X^1/Y^0 \cdot X^0 \leq x(X^1, X^0) \leq Y^1 \cdot X^1/Y^1 \cdot X^0 = \pi_q$$

Of course, in the nonhomothetic case we lose one of the bounds and must be content with the following counterparts to (8.4), involving inequality reversals

(9.8) $y(Y^1, Y^0; Y^1) \leq \pi_y$

(9.8′) $y(Y^1, Y^0; Y^0) \geq \lambda_y$

As before, various approximate interpolations to unknown $y(Y)$ or $x(X)$ functions can be made on the basis of observed (Y^i, X^i) pairs, which are subject to the duality partial derivative relations like those of (6.3)

(9.9) $(\partial x/\partial x_i)/(\partial x/\partial x_1) = y_i/y_1 \quad (i = 2, \ldots, n)$

$(\partial y/\partial y_i)/(\partial y/\partial y_1) = x_i/x_1$

Suppose the data we observe were *simultaneously* (q_j) amounts consumed and (x_j) amounts produced in a closed economy, with producers' competitive prices (y_j) identical to those facing consumers (p_j): i.e., suppose $X \equiv Q$, $Y \equiv P$. Then (9.7)'s reversal of our Paasche-Laspeyres inequalities of (7.4) would at first glance seem to threaten us with contradictions of the following type

$$a \leq z \leq b \quad \text{and} \quad b \leq z \leq a$$

Actually, such simultaneous inequalities imply no contradiction but only the equality

$$a = z = b$$

As applied to $q(Q)$, $x(X)$, $p(P)$, and $y(Y)$, this implies that the only admissible changes under our straightjacket of assumptions would be balanced changes in the ($Q = X$ and $P = Y$) variables. This is balanced growth with a vengeance—not surprising when homothetic tastes encounter homothetic or neutral technical change. With ($Q^1 = \alpha Q^0$ and $P^1 = \beta P^0$), neither the $x(\ )$ nor the $q(\ )$ functions can be identified, because there is no scatter. However,

(9.10) $x(\alpha X^0)/x(X^0) \equiv \alpha \equiv q(\alpha Q^0)/q(Q^0)$,

$y(\beta Y^0)/y(Y^0) \equiv \beta \equiv p(\beta P^0)/p(P^0)$,

a wholly natural consistency.

In many cases, as in a comparison of the production potentialities of two regions, we do not know that the x parameter uniformly shifts the frontier *out*ward. This is what the investigator wants to infer. Hence, he cannot assume that $\partial f/\partial x$ is one-signed in (9.1). But, on the supposition

that we deal with stable competitive[16] prices, with the frontiers always being concave, we can infer an outward or inward shift of X^1's frontier near the point X^0 from the index number comparison

$$(9.11) \qquad x(X^1, X^0; Y^0) \lesseqgtr 1$$

One-sided inequalities, like those of (7.4′), will be valid in the nonhomothetic case, but of course with signs reversed:

$$(9.12) \qquad x(X^1, X^0; Y^0) \geq \lambda_q$$

$$(9.12') \qquad x(X^1, X^0; Y^1) \leq \pi_q$$

If $\lambda_q > 1$, (9.12) tells us that $x(X^1, X^0; Y^0) > 1$ and X^1's frontier lies outside X^0. Alternatively, if $\pi_q < 1$, then $x(X^1, X^0; Y^1) < 1$ and we can infer that X^0's frontier lies outside X^1. But, if we are given only (X^i, Y^i) data and $\lambda_q < 1$ and $\pi_q > 1$, no qualitative inference is possible about any shiftings. Since frontiers, unlike indifference contours, can twist in the general nonhomothetic case so as to intersect, we cannot infer anything about π_q from $\lambda_q > 1$, or anything about λ_q from $\pi_q < 1$.

It could happen that *both* situations are simultaneously better than each other in this sense, that there has been a *twist* of the frontiers, as revealed by

$$(9.13) \qquad x(X^1, X^0; Y^0) \geq \lambda_q \geq 1 > \pi_q \geq x(X^1, X^0; Y^1)$$

This implication of inequality of $x(X^1, X^0; Y^0)$ and $x(X^1, X^0; Y^1)$, the former exceeding unity and the latter falling short of unity, would be impossible in the homothetic case (where twists of the frontiers are in any case ruled out); but, as mentioned, it is not surprising in the nonhomothetic case where frontiers may cross, in contrast to the case of noncrossing indifference contours.

We must warn that it is never possible, on the basis of $[Y^1, X^1; Y^0, X^0]$ observations alone, to infer that the one frontier lies *inside* the other in the neighborhood of the latter's observed point. This is because the one-sided inequality of (9.12) then becomes

$$(9.14) \qquad 1 > \lambda_q < x(X^1, X^0; Y^0) \lesseqgtr 1$$

We leave this brief survey with the remark that Moorsteen and Bergson often work, not with our $x(X^1, X^0; Y^\alpha)$ index numbers, but rather with the following variant concept:[17]

[16] If technology cannot be "lost" or "forgotten," and if X^0 precedes X^1 in time, then the fact that the earlier options are still available means that the frontier cannot shift inward: at worst one can take convex combinations of points on both frontiers, so that the new concave frontier, if anything, always shifts outward. This, however, assumes that the X vector includes (possibly as negative items) the stock of inputs; if any of the latter are omitted, then reduction of their stock in the background could shift the observed frontier inward. Also, in interregional cross-sectional comparison, as between *U.S.* and *USSR* production potentialities, we cannot assume the same effective laws of technology at both X^0 and X^1; so X^1 need not dominate X^0, and twists of the frontier are admissible.

[17] As mentioned in fn. 4, the Pollak quantity index of demand resembles this Moorsteen-Bergson variant to our $x(X^1, X^0; Y^\alpha)$. Pollak defines his q index by solving $u(\lambda^1 Q^\alpha) = u(Q^1)$, $u(\lambda^0 Q^\alpha) = u(Q^0)$ for the ratio λ^1/λ^0 as a function of $(Q^1, Q^0; Q^\alpha)$. This fails (v*)'s factor-reversal test but passes the quantity version of (i*). In the homothetic case, Pollak's $q(\ )$ agrees with our $q(Q^1, Q^0, P^\alpha) = q(Q^1)/q(Q^0)$. It might be pointed out that a price index, different from our $p\ (P^1, P^0; Q^\alpha)$, could be defined by a logic dual to that of Pollak's here: Define $p\{P^1, P^0; P^2\}$ or λ^1/λ^0 where $u^*(P^2\lambda^1) = u^*(P^1)$, $u^*(P^2\lambda^0) = u^*(P^0)$. For homothetic demand *only* will $p\{P^1, P^0; P^2\} \equiv p(P^1, P^0; Q^\alpha)$ and will $p\{,\ ;\}$ satisfy Fisher's original (i*) requiring $p\{\lambda P^0, P^0; P^2\} \equiv \lambda$. [Added in proof: Hicks, ch. 19, noting that $\lambda_q > \pi_p$ in the homothetic case and the case where Q^0 and Q^1 are indifferent, and attributing this to a "substitution effect," writes $\lambda_q - \pi_q$ in the nonhomothetic case where Q^1 and Q^0 are not indifferent, as the sum of such a substitute effect and a so-called "income effect." However, even if no good is inferior, when demand is nonhomothetic, there will always be somewhere southwest of any Q^0 point, an infinity of Q^1 points for which $\pi_q > L_q$. Our attention has been called belatedly to the useful analysis of Nissan Liviatan and Don Patinkin, which among many other points notices the importance of nonhomotheticity. Also to Richard Geary, who attempts to organize the (Q^k, P^k) for N countries where prices are in local currencies by defining 1) certain average world prices, $\bar{p}_i$ in our notation and 2) standardized exchange rate parities for the α country in terms of the hypo-

"What *multiple* of X^1 could have been produced on X^0's frontier? Call it $m(X^1, X^0)^{-1}$. Likewise define $n(X^0, X^1)$ as the *multiple* of X^0 producible on X^1's frontier."

These two last measures, m and n, can be shown necessarily to equal each other only in the homothetic case, and then they will equal our $x(X^1)/x(X^0)=x(X^1, X^0; P)$; but $m(\ ,\ )$, $n(\ ,\ )$, and $x(\ ,\ ;\ )$ need not agree in the general, nonhomothetic case. Of course, depending upon whether or not $m(X^1, X^0)\lesseqgtr 1$, X^0 lies outside, on, or inside X^1's frontier. In general, it is not possible to say that the $x(X^1, X^0; Y^\alpha)$ concept is more or less useful than the $m(X^1, X^0)$ and $n(X^0, X^1)$ concept; exactly the same one-sided bounds applies, in the nonhomothetic case, to $m(X^1, X^0)$ as $x(X^1, X^0; Y^0)$ but their magnitudes relative to each other are ambiguous.

$$(9.15)\quad \lambda_q\le x(X^1, X^0; Y^0) \gtrless m(X^1, X^0)\ge\lambda_q$$

$$\pi_q\ge x(X^0, X^1; Y^1) \gtrless n(X, X^1)\ge\pi_q^{-1}$$

Only in the homothetic frontier case do we get the two-sided identities

$$(9.16)\quad \pi_q \ge n(X^0, X^1) = x(X^1)/x(X^0) = m(X^1, X^0) \ge \lambda_q$$

To the extent that the production possibility frontier is almost linear with high elasticity of substitution, λ_q approximates $m(X^1, X^0)$ better than it does $x(X^0, X^1; Y^0)$; to the extent that the frontier has low elasticity of substitution, λ_q approximates $x(X^0, X^1; Y^0)$ better than it does $m(X^1, X^0)$.

Finally, we mention the "Gerschenkron effect," namely that when a society's frontier is augmented more in terms of one good (say, machines) than another (say, bread), the quantity mix of demand allegedly moves to favor the most-augmented good whereas its price ratio moves against that good. This common sense empirical likelihood would be a deductive necessity if tastes for the two goods were homothetic and with normal curvature. For then, from the double-limit theorem of homothetic tastes in (7.3), we can infer

$$(9.17)\quad \pi_q \le q(X^1)/q(X^0) \le \lambda_q$$

Hence, we have deduced the well-known phenomenon, noted by Kravis-Gilbert and numerous writers such as Samuelson (1974b).

SERENDIPITY THEOREM: *Each region tends empirically to fare better in terms of a comparison involving its own prices and mix, provided*

$$(9.18)\quad \{(x_2/x_1)^1 - (x_2/x_1)^0\}\{(p_2/p_1)^1 - (p_2/p_1)^0\} < 0$$

$$(9.19)\quad x(X^1, X^0; P^1) < \pi_q < \lambda_q < x(X^1, X^0; P^0)$$

$$(9.20)\quad n(X^0, X^1)^{-1} < \pi_q < \lambda_q < m(X^1, X^0)$$

Again, there is no advantage or disadvantage in the Moorsteen-Bergson variant of (9.20) over (9.19)'s conventional production index numbers. The proof of $\pi_q<\lambda_q$ in the Gerschenkron case follows immediately by arithmetic from the assumed fact that the $[(p_2/p_1), (x_2/x_1)]$ have a negative Pearsonian correlation coeffi-

---

thetical world currency, E^α, so that

$$\bar{P}=[\bar{p}_j]=\left[\sum_{\alpha=1}^{N} E^\alpha p_j^\alpha q_j^\alpha \Big/ \sum_{\alpha=1}^{n} q_j^\alpha\right]\quad (j=1,\ldots,n)$$

$$E^\alpha=\sum_{j=1}^{n}\bar{p}_j q_j^\alpha \Big/ \sum_{j=1}^{n} p_j^\alpha q_j^\alpha\quad (\alpha=1,\ldots,N)$$

Then E^α is applied as a deflator to $P^\alpha\cdot Q^\alpha$ to get a measure of real income $E^\alpha P^\alpha\cdot Q^\alpha=\bar{P}\cdot Q^\alpha$. In the easiest cases where we know the exact answer, as in the homothetic case, $n=2$ and α takes on a continuum with $p_2^\alpha/p_1^\alpha=\pi(\alpha)$, $q_2^\alpha/q_1^\alpha=\theta(\alpha)$, π and θ being observable functions, the Geary measure does *not* give the correct $p(P^\alpha)/p(P^\beta)$ and $q(Q^\alpha)/q(Q^\beta)$ even though these correct functions are uniquely inferrable. So Geary's method seems to be a throwback to the Fisher-Pearson-Frickey mechanical methods. Like them, it may turn out to have some heuristic merits for approximating useful results.

cient between the two situations (reminding one of the old-fashioned discussions of "weight-bias").

X. Concluding Warning

Empirical experience is abundant that the Santa Claus hypothesis of homotheticity in tastes and in technical change is quite unrealistic. Therefore, we must not be bemused by the undoubted elegances and richness of the homothetic theory. Nor should we shoot the honest theorist who points out to us the unavoidable truth that in nonhomothetic cases of realistic life, one must not expect to be able to make the naive measurements that untutored common sense always longs for; we must accept the sad facts of life, and be grateful for the more complicated procedures economic theory devises.

REFERENCES

S. N. Afriat, "The Theory of International Comparisons of Real Incomes and Prices," in J. D. Daly, ed., *International Comparisons of Prices and Output*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 37, New York 1972, 15–94.

A. Bergson, *National Income of the Soviet Union since 1928*, Cambridge 1961.

P. W. Bridgman, *Dimensional Analysis*, rev. ed., New Haven 1931.

F. Divisia, *Économie Rationnelle*, Paris 1928.

F. Fisher and K. Shell, *The Economic Theory of Price Indices*, New York 1972.

I. Fisher, *Purchasing Power of Money*, New York 1911.

———, *The Making of Index Numbers*, Cambridge 1922.

R. Frisch, "Necessary and Sufficient Conditions Regarding the Form of an Index Number Which Shall Meet Certain of Fisher's Tests," *J. Amer. Statist. Assn.*, Dec. 1930, *25*, 297–406.

———, *New Methods of Measuring Marginal Utility*, Tübingen 1932.

———, "Annual Survey of General Economic Theory: The Problem of Index Numbers," *Econometrica*, Jan. 1936, *4*, 1–38.

R. C. Geary, "A Note on the Comparison of Exchange Rates and Purchasing Power Between Countries," *J. Royal Statist. Soc.*, Part I, 1958, *121*, 97–99.

A. Gerschenkron, *A Dollar Index of Soviet Machinery Output*, Santa Monica 1951.

J. R. Hicks, *A Revision of Demand Theory*, Oxford 1956.

H. Hotelling, "Edgeworth's Taxation Paradox and the Nature of Demand and Supply Functions," *J. Polit. Econ.*, Oct. 1932, *40*, 577–616.

H. Houthakker, "Additive Preferences," *Econometrica*, Apr. 1960, *28*, 244–57; "Errata," *Econometrica*, July 1962, *30*, 633.

———, "Some Problems in the International Comparison of Consumption Patterns," in *L'Evaluation et le Rôle des Besoins de Consommation dans les Divers Régimes Economiques*, Paris 1963.

D. M. Iklé, "A New Approach to Index Numbers," *Quart. J. Econ.*, May 1972, *86*, 188–211.

D. Jorgenson and Z. Griliches, "The Explanation of Productivity Change," *Rev. Econ. Stud.*, July 1964, *34*, 249–82.

W. W. Leontief, "Composite Commodities and the Problem of Index Numbers," *Econometrica*, Jan. 1936, *4*, 39–59.

N. Liviatan and D. Patinkin, "On the Economic Theory of Price Indexes," *J. Econ. Develop. Cult. Exchange*, Apr. 1961, *9*, 502–36.

D. McFadden, "Constant Elasticity of Substitution Production Function," *Rev. Econ. Stud.*, June 1963, *30*, 73–83.

L. W. McKenzie, "Demand Theory Without a Utility Index," *Rev. Econ. Stud.*, June 1957, *24*, 185–89.

R. H. Moorsteen, "On Measuring Productive Potential and Relative Efficiency," *Quart. J. Econ.*, Aug. 1961, *75*, 451–67.

R. A. Pollak, "The Theory of the Cost of Living Index," unpublished paper, Univ. Pennsylvania, June 1971.

R. Roy, *De l'Utilité; Contribution à la Théorie des Choix*, Paris 1942.

R. Ruggles, "Price Indexes and International

Price Comparisons," in *Ten Economic Studies in the Tradition of Irving Fisher*, New York 1967.

P. A. Samuelson, *Foundations of Economic Analysis*, Cambridge 1947.

———, "Consumption Theory in Terms of Revealed Preference," *Economica*, Nov. 1948, *15*, 243–58; reprinted in J. E. Stiglitz, ed., *The Collected Scientific Papers of Paul A. Samuelson*, Vol. I, Cambridge 1966, 64–74.

———, "Evaluation of Real National Income," *Oxford Econ. Pap.*, Jan. 1950, *2*, 1–29; reprinted in J. E. Stiglitz, ed., *The Collected Scientific Papers of Paul A. Samuelson*, Vol. II, Cambridge 1966, 1044–72.

———, "Consumption Theorems in Terms of Overcompensation Rather Than Indifference Comparisons," *Economica*, Feb. 1953, *20*, 1–9.

———, "Parable and Realism in Capital Theory: The Surrogate Production Function," *Rev. Econ. Stud.*, June 1962, *29*, 193–206; reprinted in J. E. Stiglitz, ed., *The Collected Scientific Papers of Paul A. Samuelson*, Vol. I, Cambridge 1966, 325–38.

———, "Using Full Duality to Show that Simultaneously Additive Direct and Indirect Utilities Implies Unitary Price Elasticity of Demand," *Econometrica*, Oct. 1965, *33*, 781–96; reprinted in R. C. Merton, ed., *The Collected Scientific Papers of Paul A. Samuelson*, Vol. III, Cambridge 1972, 71–86.

———, "Unification Theorem for the Two Basic Dualities of Homothetic Demand Theory," *Proc. Nat. Acad. Sci., U.S.*, Sept. 1972, *69*, 2673–74.

———, (1974a) "Remembrances of Frisch," *European Econ. Rev.*, 1974 forthcoming.

———, (1974b) "Analytical Notes on Real Income Measures," *Econ. J.*, 1974 forthcoming.

J. A. Schumpeter, *Business Cycles*, Vol. II, New York 1939, 452–58.

R. W. Shephard, *Cost and Production Functions*, Princeton 1953.

H. Staehle, "A Development of the Economic Theory of Price Index Numbers," *Rev. Econ. Stud.*, Feb. 1935, *3*, 163–88.

S. Star and R. E. Hall, "An Approximate Divisia Index of Total Factor Productivity," Univ. British Columbia department of economics disc. pap. 73–19, Oct. 1973.

S. Swamy, "Consistency of Fisher's Tests," *Econometrica*, July 1965, *33*, 619–23.

———, "On Samuelson's Conjecture," *Indian Econ. Rev.*, Oct. 1970, *5*, 169–75.

H. Theil, *Economics and Information Theory*, Chicago 1967.

M. J. Ulmer, *The Economic Theory of Cost of Living Index Numbers*, New York 1949.

H. Uzawa, "Production Functions with Constant Elasticities of Substitution," *Rev. Econ. Stud.*, Oct. 1962, *29*, 291–99.

J. Ville, "The Existence Condition of a Total Utility Function and of an Index of Price," *Rev. Econ. Stud.*, 1951, *19*, 128–32, a translation of a 1946 paper from the *Annales de l'Université de Lyon.*

A. Wald, "A New Formula for the Index of Cost of Living," *Econometrica*, Oct. 1939, *7*, 319–31.

H. Wold, *Demand Analysis*, New York 1953.

Technological Change in the Soviet Collective Farm

By RAVEENDRA N. BATRA*

Models of a cooperative enterprise have been utilized by Evsey Domar, and Walter Oi and Elizabeth Clayton to analyze the economic behavior of a Soviet *kolkhoz* or collective farm. Its members are assumed to be interested in maximizing per capita earnings from the cultivation of a given supply of land which is provided by the state in payment for a fixed amount of rent. The crops produced by the *kolkhoz* are assumed to be sold at prices determined not by forces of demand and supply but by the planners who are part of the Soviet planning hierarchy. Several policy devices (such as exogenous changes in product prices and rent) suggest themselves from these studies—devices which not merely satisfy academic interest but which at one time or another have been introduced by the Soviet authorities to boost the output levels in agriculture.

These theoretical models, however, suffer from two important omissions. First, they ignore the implications of technological change on the level and the pattern of agricultural output of the *kolkhoz*; and second, they say nothing about the consumption side of the collective farm.

The purpose of this paper is to fill this gap. The significance of our analysis can be gauged from the facts that (i) despite its recent failures and relative inefficiency in comparison with *U.S.* agriculture, there is no denying that the Soviet agriculture has experienced a good deal of technological change,[1] and (ii) the real incomes of the *kolkhoz* members are determined not only by how much they produce and the prices they receive, but also by the prices and the quantum of the goods they consume. The basic structure of Soviet agriculture is based on two types of collective enterprises: the state farm (*sovkhoz*), which is owned by the state and operated as any rural factory with hired managers and workers, and the collective farm (*kolkhoz*), which is similar to a producers' cooperative and where peasants share in the net income of the farm. As with the analyses by Domar and Oi and Clayton, our concern here is with the collective and not the state farms.

I. The Formulation of the Problem

The Soviet *kolkhoz* is typically engaged in the production of two different types of crops: One part of the land is devoted to collective cultivation to produce the collective crop X, and the other part is divided among the members to produce what may be called the private crop Y. The peasant

* Professor of economics, Southern Methodist University. I wish to express my gratitude to Gerhard Prosi, William Russell, and Paul Heyne for useful discussions. Thanks are also due to the referee for several suggestions which led to improvements in the exposition.

[1] Technological change is normally defined in terms of an increase in the productivity of one or all of the existing factors of production. Accepting D. Gale Johnson's estimates, p. 233, the output-input ratio in Soviet agriculture increased by 10 to 20 percent between 1938 and 1950 and by about 30 percent between 1950 and 1959. There is thus unequivocal evidence of technological change in Soviet agriculture. It is, of course, necessary to remind ourselves that between 1928 and 1938 there was an unprecedented decline in the output-input ratio in agriculture due primarily to collectivization and the consequent disorganization of rural life. Although the results in this paper are presented in terms of technological change, our model is quite capable of handling the implications of exogenous declines in productivity, because the effects of the latter are simply the opposite of the effects of technical progress.

on the *kolkhoz* thus earns his income from the sale of X and Y at fixed prices, p_x and p_y. Following Oi and Clayton, we assume that all peasants receive equal amounts of land and equally share the payment of rent R or the lump sum tax. Let N be the total supply of labor, N_i the input of labor devoted to the ith crop ($i=x, y$), T the amount of land available to the *kolkhoz*, and T_i the input of land devoted to the production of the ith crop. Then, if each member of the *kolkhoz* contributes the same number of hours of labor, it is clear that the size of T_y and hence T_x is proportional to the size of N and hence to the number of members. Symbolically, we may write

$$N_x + N_y = N \tag{1}$$

$$T_x + T_y = T \tag{2}$$

$$T_y = kN \tag{3}$$

where $k>0$ is the constant rate at which land is transferred to private plots as a result of the increase in the membership.[2] The per capita income of the *kolkhoz* is given by

$$w = \frac{p_x X + p_y Y - R}{N} \tag{4}$$

where $X=F_x(T_x, N_x)$ and $Y=F_y(T_y, N_y)$ are the production functions. We assume that the production functions are homogeneous of degree one, so that we can write

$$X = F_x(T_x, N_x) = T_x F_x(1, N_x/T_x) = T_x f_x(\rho_x) \tag{5}$$

$$Y = F_y(T_y, N_y) = T_y F_y(1, N_y/T_y) = kNf_y(\rho_y) \tag{6}$$

where f_i equals the average productivity of land and ρ_i denotes the labor-land ratio in the ith crop. We further assume that each *kolkhoz* complex seeks to maximize its per capita welfare represented by a social utility function,

$$u = u(c_x, c_y, c_m) \tag{7}$$

where u stands for per capita utility, and c_x, c_y, and c_m, respectively, denote the per capita consumption of goods, X, Y, and M, and where M stands for the bundle of industrial goods which are produced in the nonfarm sector of the Soviet economy.[3] In other words, the real income of each peasant depends on his consumption bundle of the two crops and a manufactured good which the *kolkhoz* imports from the nonagricultural sector of the Soviet economy at a given price, p_m, in exchange for the export of X and Y. In what follows, the per capita variables will be denoted by the lower case letters. Thus, $x=X/N$, $c_x=C_x/N$, and so on. In view of equation (4), the budget constraint on the *kolkhoz* is provided by

$$\begin{aligned} w &= p_x x + p_y y - r \\ &= p_x c_x + p_y c_y + p_m c_m \end{aligned} \tag{8}$$

With this last equation, the production as well as the consumption sides of our model of the Soviet *kolkhoz* are complete.

The problem may be formulated as follows: The *kolkhoz* members seek to maximize per capita utility subject to the budget constraint furnished by equation (8). This may be accomplished by forming the Lagrangian L:

[2] Oi and Clayton, p. 38, have provided the following explanation for the fact that k is an institutionally determined constant. Let each member receive t hectares for his private plot, contribute n man-days of labor per year, and let M be the number of members of the *kolkhoz*. Then, labor supply N equals nM and the hectares of land devoted to private plots equal tM or $(t/n)N$. Then $k=t/n$ is the fixed rate determining the diversion of land into private plots as M or N rises.

[3] There is considerable evidence to suggest that during the Stalin era, prices of farm goods were set at low levels and prices of manufactured goods consumed by the farm population were set at very high levels in order to achieve high rates of investible surplus. As a consequence, the peasant real incomes were abysmally low. With the help of the model developed in this paper, it is possible to analyze theoretically the implications for farm incomes of pricing policies followed in the Stalin era.

$$(9)\quad L=u(c_x, c_y, c_m) + \lambda - \left(\frac{p_xX+p_yY-R}{N}-p_xc_x-p_yc_y-p_mc_m\right)$$

where $\lambda>0$. The first-order conditions for an interior maximum can be obtained by differentiating equation (9) partially with respect to c_i, X, and N, so that

$$(10)\quad u_i - \lambda p_i = 0 \qquad (i = x, y, m)$$

$$(11)\quad \frac{\lambda}{N}\left(p_x + p_y \frac{\partial Y}{\partial X}\right) = 0$$

$$(12)\quad \frac{\lambda}{N}\left[p_x \frac{\partial X}{\partial N} + p_y \frac{\partial Y}{\partial N} - \frac{p_xX + p_yY - R}{N}\right] = 0$$

where $u_i=\partial u/\partial c_i>0$ is the marginal utility of the ith good. From equation (10),

$$(13)\quad \frac{u_x}{u_y} = \frac{p_x}{p_y} \quad \text{and} \quad \frac{u_m}{u_y} = \frac{p_m}{p_y}$$

which implies that the marginal rate of substitution between any two commodities equals the ratio between their prices. From equation (11), since λ/N is nonzero,

$$(14a)\quad -\frac{\partial Y}{\partial X} = \frac{p_x}{p_y}$$

which implies that the marginal rate of transformation between the two crops equals the ratio between their prices. Actually, equation (14a) represents the equilibrium condition for resource allocation in the short run where the membership and hence N are fixed. This can be appreciated by remembering that in deriving equation (11), equation (9) was differentiated *partially* with respect to X while N was kept constant. Condition (14a) can be alternatively stated as

$$(14b)\quad p_xf_x' = p_yf_y'$$

where $f_i'=df_i/d\rho_i$ is the marginal product of labor in the ith crop.[4] Equation (14b) suggests that the *kolkhoz* allocates the fixed supply of labor in such a way that the value of the marginal product of labor in the two crops is equated.

Finally, in the long run, N becomes a variable, because the *kolkhoz* can reduce or increase the labor input by varying the membership. The long-run equilibrium condition, furnished by equation (12) can be simplified to

$$(15)\quad p_x(f_x - \rho_xf_x') = \frac{R}{T}$$

where $f_i-\rho_if_i'$ is the marginal product of land.[5] Equation (15) suggests that in the long run, the value of the marginal productivity of land in the collective crop is equated to the total amount of rent per unit of land.

II. Rent, the Terms of Trade, and the Peasant Welfare

The analysis presented above is a necessary prelude to examining the implications of a change in the policy parameters, p_x, p_y, p_m, and R on the real income or welfare of each individual member of the *kolkhoz*.

[4] Differentiating the two production functions and dividing, we obtain

$$\frac{dY}{dX} = \frac{f_y'dN_y + (f_y - \rho_yf_y')dT_y}{f_x'dN_x + (f_x - \rho_xf_x')dT_x}$$

Since $dT_x=dT_y=0$ and $dN_x=-dN_y$ in the short run where N is given, $dY/dX=-f_y'/f_x'=-p_x/p_y$.

[5] The production functions given by equations (5) and (6) can be written as $X=F_x(N_x, T-kN)$ and $Y=F_y(N-N_x, kN)$, so that in view of Euler's theorem, equation (12) may be written as

$$p_x\left[f_x\frac{\partial N_x}{\partial N} - k(f_x - \rho_xf_x')\right] + p_y\left[f_y'\left(1-\frac{\partial N_x}{\partial N}\right) + k(f_y - \rho_yf_y')\right] - \frac{1}{N}\{p_x[f_x'N_x + (f_x - \rho_xf_x')(T-kN)] + p_y[f_y'(N-N_x) + (f_y - \rho_yf_y')kN] - R\} = 0$$

This in turn simplifies to

$$R - Tp_x(f_x - \rho_xf_x') = 0$$

What turns out to be most interesting is that a change in any one of the three prices affects the peasant welfare in different ways, either qualitatively or quantitatively.

Differentiating the utility function given by equation (7) totally with respect to p_x alone and using equation (13), we have

$$(16)\quad \frac{1}{u_m}\frac{du}{dp_x} = \frac{p_x}{p_m}\cdot\frac{dc_x}{dp_x} + \frac{p_y}{p_m}\cdot\frac{dc_y}{dp_x} + \frac{dc_m}{dp_x}$$

Differentiation of the budget constraint furnished by equation (8) yields

$$(17)\quad \frac{dw}{dp_x} = c_x + p_x\frac{dc_x}{dp_x} + p_y\frac{dc_y}{dp_x} + p_m\frac{dc_m}{dp_x}$$

Let the change in per capita utility, du/u_m, represent the change in per capita real income of the *kolkhoz* measured in terms of commodity M. Call this dz. Substitution of equation (17) in equation (16) then furnishes

$$(18)\quad \frac{dz}{dp_x} = \frac{1}{p_m}\left[\frac{dw}{dp_x} - c_x\right]$$

In the same way, we can derive

$$(19)\quad \frac{dz}{dp_y} = \frac{1}{p_m}\left[\frac{dw}{dp_y} - c_y\right] \quad \text{and}$$

$$\frac{dz}{dp_m} = \frac{1}{p_m}\left[\frac{dw}{dp_m} - c_m\right]$$

The next step in our analysis calls for obtaining the expression for dw/dp_i. Using equations (14) and (15) and Euler's theorem, equation (4) becomes

$$(20)\quad w = p_yf_y' + k[p_y(f_y - \rho_yf_y') - p_x(f_x - \rho_xf_x')]$$

The total differentiation of equations (14) and (15) yields:

$$(21)\quad d\rho_x = \frac{(f_x - \rho_xf_x')dp_x - dR/T}{p_x\rho_xf_x''}$$

$$(22)\quad d\rho_y = \frac{f_xdp_x - \rho_xf_y'\,dp_y - dR/T}{p_y\rho_xf_y''}$$

where $f_i'' = df_i'/d\rho_i < 0$ for linear homogeneous and concave production functions.

Differentiating equation (20) with respect to p_x, p_y, and R, and using equations (21) and (22), we obtain

$$(23)\quad \frac{dw}{dp_x} = \frac{f_x}{\rho_x}, \qquad \frac{dw}{dp_y} = kf_y \quad \text{and}$$

$$\frac{dw}{dR} = -\frac{(1 - k\rho_y) + k\rho_x}{T\rho_x}$$

These equations reveal that a rise in p_x or p_y leads to a rise in w. On the other hand, a decline in R also causes a rise in w, because $1 - k\rho_y = N_x/N > 0$. One may also conclude that a rise in the availability of land will have exactly the same effect on w as a decline in R.[6]

Substituting equations (23) in equations (18)–(20), we obtain

$$(24)\quad \frac{dz}{dp_x} = \frac{1}{p_m}\left[\frac{f_x}{\rho_x} - c_x\right] = \frac{(X - C_x) + N_yc_x}{N_xp_m} = \frac{1}{p_m}\left[\frac{N}{N_x}e_x + \frac{N_y}{N_x}c_x\right]$$

$$(25)\quad \frac{dz}{dp_y} = \frac{1}{p_m}[kf_y - c_y] = \frac{Y - C_y}{p_mN} = \frac{e_y}{p_m}$$

$$(26)\quad \frac{dz}{dp_m} = -\frac{c_m}{p_m}$$

where e_i denotes the per capita export of the ith crop ($i = x, y$). In obtaining equation (26), dw/dp_m has been equated to zero because a change in p_m leaves w unchanged. A close examination of the last three equations reveals that a change in the three prices affects the peasant real in-

[6] In the case where $dR = 0$ and $dT > 0$, equation (26) becomes

$$dw/dT = R(1 - k\rho_y + k\rho_x)/\rho_xT^2$$

come in a different manner. Whereas an increase in p_y raises the real income in proportion to the initial quantum of the per capita export of Y, the increase in p_x causes a rise in the real income in a proportion greater than the initial volume of per capita exports of X. Other things remaining equal, a rise in p_x would cause a greater rise in the peasant real income than an equivalent rise in p_y. Furthermore, with the rise in p_x, the improvement in the real income would be higher, the higher is the value for N_y, and hence, the lower is the value for N_x. However, no such result is available in the case of a rise in p_y.

The effects of a change in p_m on the real income are qualitatively different from the effects of a change in either p_x or p_y. A rise in p_m causes a decline in the peasant real income in proportion to the initial volume of M's per capita imports.

Finally, as far as the effects of a change in the rent on the real income are concerned, it can be easily seen that they are identical to the expression for $(1/p_m)\cdot(dw/dR)$ which may be obtained from equations (23).

III. The Implications of Technological Change

This section is concerned with the implications of technical improvements for factor proportions in the two activities, the outputs of the two crops, and the size of the *kolkhoz* membership.[7]

Let λ_i be the agent of technical change which raises the marginal productivity of labor in the ith crop, and let β_i be the corresponding agent for land. Both λ_i and β_i are initially equal to one. The two production functions in the presence of technical advance are written as:

$$(27) \quad X = F_x(\lambda_x N_x, \beta_x T_x) = \beta_x T_x f_x(\lambda_x \rho_x/\beta_x)$$

$$(28) \quad Y = F_y(\lambda_y N_y, \beta_y T_y) = \beta_y T_y f_y(\lambda_y \rho_y/\beta_y)$$

Technical progress in the ith activity is labor saving if $d\lambda_i > 0$; it is land saving if $d\beta_i > 0$; and it is neutral if $d\beta_i = d\lambda_i = d\alpha_i > 0$. The marginal productivities are then given by $\lambda_i f_i'$ and $\beta_i f_i - \lambda_i \rho_i f_i'$. The short- and the long-run equilibrium conditions given by equations (14) and (15) are replaced by

$$(29) \quad p_x \lambda_x f_x' = p_y \lambda_y f_y'$$

$$(30) \quad p_x(\beta_x f_x - \lambda_x \rho_x f_x') = \frac{R}{T}$$

Finally, equations (1)–(3) can be written as

$$(31) \quad T_x = T - kN$$

$$(32) \quad T\rho_x = N[(1 - k\rho_y) + k\rho_x]$$

A. *The Short-Run Implications*

The short-run response of the *kolkhoz* outputs to technical change is strikingly different from their response in the long run. In the short run, the total labor input is given, so that condition (30) does not apply. The labor-land ratio in the two crops is then determined by two equations (29) and (32). For the sake of simplicity and without loss of generality, we assume that initially $p_x = p_y = 1$. Differentiating equation (32) totally and keeping N constant, we obtain

$$(33) \quad T_x d\rho_x = - T_y d\rho_y$$

In other words, in the short run where membership is constant, an increase in the labor-land ratio in one activity must be matched by a decline in labor-land ratio in the other. Differentiating equation (29) totally, using equation (33) and remembering that initially $\lambda_i = \beta_i = 1$, we obtain

$$(34) \quad d\rho_y = \frac{1}{A}\left[(f_y' + \rho_y f_y'')d\lambda_y - (f_x' + \rho_x f_x'')d\lambda_x + \rho_x f_x'' d\beta_x - \rho_y f_y'' d\beta_y\right]$$

[7] Since technological change must always raise peasant real incomes, such effects will not be discussed in the interest of brevity.

where $A=-(f_y''T_x+f_x''T_y)/T_x>0$. Since $f_i''<0$, it is clear that the implications of the labor-saving improvement in any crop (so that $d\lambda_i>0$ and $d\beta_i=0$) for ρ_i are indeterminate.

The response of factor proportions to the land-saving improvements is, however, categorical. A land-saving improvement in Y ($d\beta_i>0$ and $d\lambda_i=d\beta_x=0$) leads to a rise in ρ_y and hence to a decline in ρ_x. By contrast, $d\beta_x>0$ leads to a decline in ρ_y and to a rise in ρ_x.

The implications of neutral technical progress where $d\lambda_i=d\beta_i=d\alpha_i$ for ρ_i are also determinate. Here equation (34) reduces to

$$d\rho_y = \frac{1}{A}\left[f_y' \, d\alpha_y - f_x' \, d\alpha_x\right] \tag{35}$$

The output effects of the technical advance depend on whether or not ρ_x and ρ_y change in a categorical manner. These effects can be determined by differentiating equations (27) and (28) totally and using equations (33) and (34); however, an independent geometrical exposition is available in Figure 1 where we have adapted the traditional Edgeworth-Bowley box diagram to the case of the two-crop model of the *kolkhoz*. In Figure 1, the output of X is measured northeast from the O_x origin and that of Y southwest from the O_y origin. Since the supply of labor is given in the short run, the inputs of land in X and Y are fixed. Let RO_y be the initial amount of land employed in the production of Y. Then the slope of RB equals the inverse of k (i.e., $1/k=BO_y/RO_y$). However, labor in each crop is a variable input and its allocation determines the output levels of X and Y. Since T_i is fixed, the locus of production points is given by a vertical line such as RQ and along this locus the X and Y isoquants (like X_1 and X_2 and Y_1 and Y_2) intersect each other. This results from the fact that although labor gets optimally allocated, land being fixed in each activity

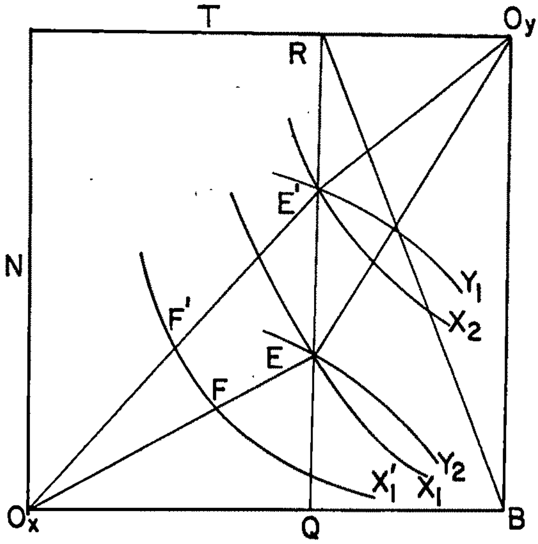


FIGURE 1

does not, so that marginal rates of factor substitution are not the same in the two activities, as is normally the case with the traditional box diagrams where all the Paretian optimality conditions of production are satisfied.

The actual production point on the locus is determined by the set of product prices and rent. Suppose E is the actual production point. Then the output of X equals EO_x and ρ_x is given by the slope of EO_x, whereas the output of Y equals EO_y and ρ_y is given by the slope of EO_y. Now suppose there occurs a neutral technical progress in X alone. Consequently, its isoquant shifts towards the origin to X_1', ρ_x remains unchanged, and the immediate effect is that the output of X rises approximately by EF. From our earlier analysis, we know that ρ_x rises and ρ_y declines in order to restore equilibrium. This means that the production point must shift vertically to say E' where ρ_x has risen and ρ_y declined. It can be observed that in the short-run equilibrium, the output of X actually rises by $E'F'$ and that of Y declines to O_yE'. One can show analogously that a neutral improvement in Y will raise

the output of Y and lower the output of X.

Proceeding in a similar way, we can show that a land-saving improvement in any crop raises the output of that crop at the expense of the output of the other; however, the effects of a labor-saving improvement, because of its uncertain effect on ρ_i, are indeterminate.

B. *The Long-Run Implications*

As suggested before, the labor input becomes variable in the long run and thus exerts an important influence on the response of outputs to technical improvements. The equilibrium conditions which determine ρ_x and ρ_y are now given by equations (29) and (30) and equation (32) comes into play only in determining the change in N. Differentiating equations (29) and (30) totally and solving them simultaneously, we obtain:

$$(36)\quad d\rho_x = -\rho_x d\lambda_x + \left[\rho_x + \frac{f_x - \rho_x f_x'}{\rho_x f_x''}\right] d\beta_x$$

$$(37)\quad d\rho_y = -\left[\rho_y + \frac{f_y'}{f_y''}\right] d\lambda_y + \rho_y d\beta_y + \frac{f_x'}{f_y''} d\lambda_x + \frac{f_x - \rho_x f_x'}{\rho_x f_y''} d\beta_x$$

In view of our short-run analysis presented above, some results are immediately striking. First of all, the technical advance in Y has no repercussions for ρ_x, simply because Y does not enter in the long-run equilibrium condition. Secondly, the implications of a labor-saving improvement in X for the labor-land ratio in both crops are unambiguous. A labor-saving improvement in X (so that $d\lambda_x > 0$) leads to a decline in ρ_x at the original level of output. The balance in the equilibrium condition (29) is maintained if f_y' rises or if ρ_y declines, which is certainly within the realm of possibility because, with variable labor input, ρ_x and ρ_y do not have to change in the opposite direction, as was the case in the short run. In fact, either ρ_x is unaltered or both ρ_x and ρ_y move in the same direction.

The implications of a labor-saving advance in Y for ρ_y, however, are uncertain. Here $d\lambda_y > 0$ leads to a decline in ρ_y, but since ρ_x and hence f_x' are unaffected, the balance in equation (29) can be maintained only by a decline in f_y' and hence a rise in ρ_y. Thus, ρ_y is subject to two conflicting pulls and a priori we do not know whether the change is positive or negative. For a similar reason, $d\beta_x > 0$ also results in ambiguous effects on ρ_x.

All other types of technical progress including the neutral improvements generate unequivocal movements in ρ_y and/or ρ_x. In the case of neutral improvements, where $d\lambda_i = d\beta_i = d\alpha_i$, equations (36) and (37) reduce to

$$d\rho_x = \frac{f_x - \rho_x f_x'}{\rho_x f_x''} d\alpha_x$$

$$d\rho_y = -\frac{f_y'}{f_y''} d\alpha_y + \frac{f_x}{\rho_x f_y''} d\alpha_x$$

The long-run effects of technical progress on the two outputs are more uncertain than the short-run effects even where ρ_y and ρ_x move unambiguously. This is because N is now variable. Differentiating equation (32), we obtain

$$(38)\qquad dN = \frac{T_x d\rho_x + T_y d\rho_y}{(1 - k\rho_y) + k\rho_x}$$

Since the denominator of equation (38) is positive, dN has the same sign as $d\rho_i$. It is clear that a neutral or a land-saving technical progress in X must cause a decline in membership in the long run because in these cases both $d\rho_x$ and $d\rho_y$ are negative. By contrast, a neutral or a land-saving technical progress in Y must cause an increase in the membership in the long run,

because here $d\rho_y > 0$ and $d\rho_x = 0$. In all other cases, the sign of dN is indeterminate.

Because of the changes in N, it turns out that only a neutral or a land-saving improvement in Y generates categorical effects on the two outputs. This can be discovered by differentiating the production functions given by equations (27) and (28) totally and using equations (36), (37), and (38). However, the expressions for dX and dY are onerously lengthy and our point can be well illustrated in terms of diagrams which also serve to underlie the role played by the variability of the labor input in generating ambiguous effects on the output of the two crops. In what follows, we will explicitly consider the two simpler cases of neutral technical progress in X or Y and summarize results for others.

Consider first the case of a neutral improvement in Y. In Figure 2, a neutral improvement in Y causes its Y_2 isoquant to shift towards the origin. The immediate effect is that its output rises approximately by GE at the original level of ρ_y given by the slope of EO_y. However, the neutral improvement in Y causes the labor input to rise in the long run, thereby shifting the

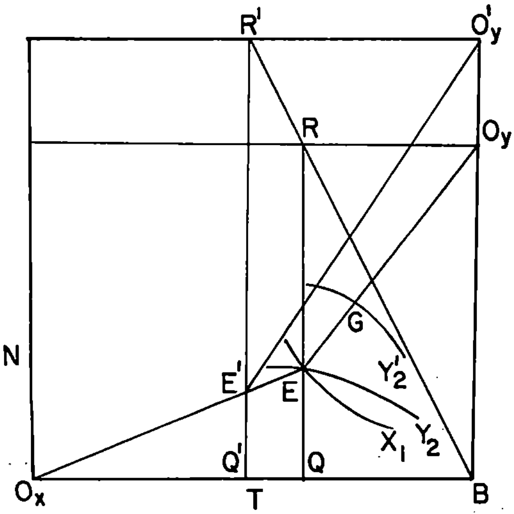


FIGURE 2

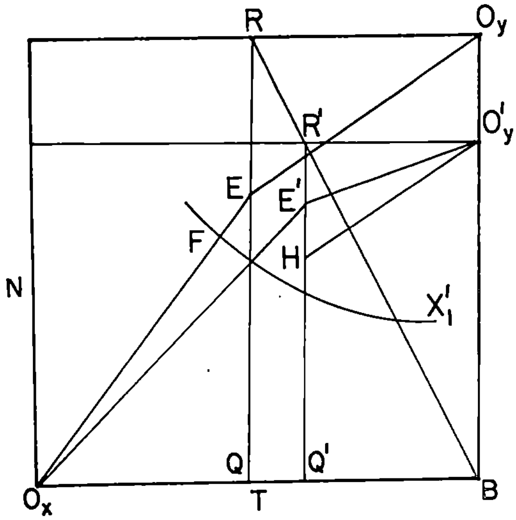


FIGURE 3a

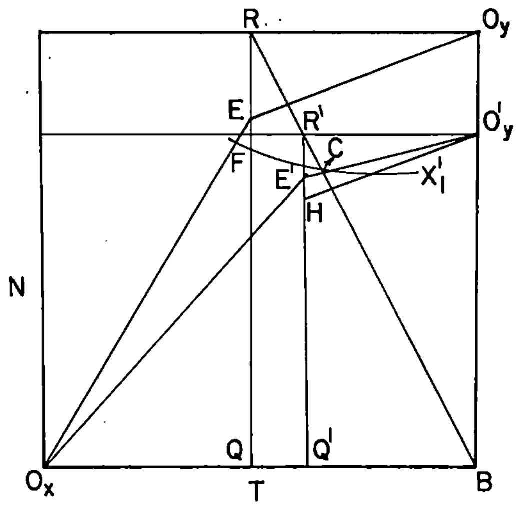


FIGURE 3b

Y origin to O_y' where O_yO_y' is the increase in labor supply. This in turn leads to the diversion of more land into the private plots. Thus, $R'O_y'$ is the amount of land employed in Y now as compared to the previous utilization of RO_y. The locus of production points also now shifts to $R'Q'$. Since ρ_x is unchanged, the new production point is given by E', showing that ρ_y has increased from the pre-improvement slope

TABLE 1

Classification of the Technical Change	The sign of			
	dN	dw	dX	dY
$d\lambda_x > 0$	Indeterminate	>0	Indeterminate	Indeterminate
$d\beta_x = 0$	<0	>0	Indeterminate	<0
$d\alpha_x = d\lambda_x = d\beta_x > 0$	<0	>0	Indeterminate	<0
$d\lambda_y > 0$	Indeterminate	>0	Indeterminate	Indeterminate
$d\beta_y > 0$	>0	>0	<0	>0
$d\alpha_y = d\lambda_y = d\beta_y > 0$	>0	<0	<0	>0

of EO_y to that of $E'O_y'$. Thus, neutral technical progress in Y leads to a rise in its output at the expense of the output of X.

The effects of a neutral improvement in X are examined in Figure 3. First, consider Figure 3a where the original X and Y isoquants are not drawn in order to avoid cluttering up the diagram. As the X isoquant moves backwards to X_1', the immediate effect again is the rise in the output of X approximately by EF at the original level of ρ_x. However, the neutral improvement in X causes the membership to decline in the long run. Consequently, the Y origin shifts vertically downwards to O_y', thereby showing that the labor input has declined by O_yO_y'. The locus of production points now shifts to $R'Q'$. To find the new production point, draw HO_y' parallel to EO_y. Now since ρ_y has declined as a result of the neutral improvement in X, the new production point must lie on $R'Q'$ somewhere above the point H. Such a production point in Figure 3a is given by E', which shows that the output of X has risen further and the output of Y has declined.

However, if the decline in membership is very large, the output of X may actually decline in the final equilibrium. Such a possibility is demonstrated in Figure 3b where the new output of X, given by $E'O_x$, is below its original output level, which equals CO_x or FO_x. What is of greater interest here is that the output of Y has also declined. This appears paradoxical, but any semblance of paradox disappears when it is remembered that with a considerably smaller input of labor, it is conceivable that both outputs may decline even though the technology in X is now better than before. By using similar diagrams, the reader can analyze the long-run implications of the other types of technical change for the two output levels. Table 1 summarizes our results.

IV. Conclusions and Some Policy Implications

A good deal of technical change has occurred in the Soviet economy since the 1917 revolution. The main focus of this paper is on the theoretical implications of different types of technical improvements in one sector of the Soviet economy—the Soviet collective farm—for the pattern and level of output in agriculture.

Our analysis shows that the way collective farms are organized, technical progress in the production of collective crops may actually lead in the long run to a decline in the output of both crops produced by the *kolkhoz*, thanks in large measure to *kolkhoz* members who seek to and are allowed to maximize per capita real income. In general, technical progress in Y raises its output, and if the improvement in Y is neutral or land saving, the output of Y rises at the expense of the output of X. In all cases of technical progress, the per capita real income of the *kolkhoz* must rise. We have not shown this explicitly, but the interested reader can readily derive the

relevant algebraic expressions from the foregoing technique.

The basic agricultural objectives of the Soviet planner have always been the substantial increase in agricultural output and the elimination of the class differentials and distinctions in terms of annual incomes between the rural peasant and the urban industrial proletariat. Before 1950, the emphasis in achieving these goals was upon extensive methods of cultivation, but since the end of the Stalin era, the stress is on intensive methods through increased investment in agriculture and application of superior techniques. The present analysis suggests that in order to raise the agricultural output of both crops and the living standards of the peasants, the Soviet planner should encourage the collective farms to introduce more efficient techniques in the production of both crops. Quite often productivity has been increasing in the production of private crops, whereas collective crops have lagged behind. This works for the benefit of the peasants but is detrimental to the interests of the rest of the Soviet society, especially when the bulk of the collective crops is procured by the state for consumption by the nonfarm community. Nevertheless a disproportionate emphasis in introducing new techniques in the collective crop is not desirable. In the long run this may lead to a reduction in membership, thereby precluding the full realization of benefits, in terms of the two outputs, of the introduction of the advanced techniques.

The second section of the paper was concerned with the effects of a change in rent as well as the prices of the goods consumed and produced by the peasants on their real incomes. The most interesting result was that, other things being equal, a rise in the price of collective crop would raise the peasant real income more than a corresponding rise in the price of the private crop, and vice versa.

REFERENCES

E. D. Domar, "The Soviet Collective Farm as a Producer Cooperative," *Amer. Econ. Rev.*, Sept. 1966, *56*, 734–57.

D. G. Johnson, "Agriculture Production," in A. Bergson and S. Kuznets, eds., *Economic Trends in the Soviet Union*, Cambridge 1963, ch. 5.

W. Y. Oi and E. M. Clayton, "A Peasant's View of a Soviet Collective Farm," *Amer. Econ. Rev.*, Mar. 1968, *58*, 37–59.

Investments in Human Capital and Growth in Personal Income 1956-1966

By THOMAS JOHNSON AND FREDERICK J. HEBEIN*

Economists have accepted the view of personal nonproperty incomes as the returns to the quantity of human capital which the individual possesses. Numerous studies have estimated internal rates of return or rental rates to investments in human capital.[1] Several studies have used the basic human capital model in estimating the contribution of education to the growth in national income,[2] while others have applied the concept to the analysis of the distribution of income.[3] However, in the empirical estimation of parameters, most of these studies of growth and income distribution use either a single cross-section of data or else make separate estimates for each of a series of cross-sections.

The main purpose of this paper is to estimate the rate of growth in personal income in the United States using a series of cross-sections in a single regression. A second purpose is to explore the effects of including a finite life correction in Johnson's earlier model, thus estimating internal rates of return rather than rental rates. Finally we examine the effects of alternative specifications of the on-the-job training (*OJT*) function.[4] In general, these corrections are found to be minimal.

After the biasing effects of exogenous growth is removed, the estimates of rates of return, depreciation, and *OJT* investment are in reasonable agreement with previous estimates. Moreover, the rates of depreciation are more reassuring than the very high rates previously estimated by Johnson (1969, 1970). The lifetime incomes by cohort implied by the parameter estimates exhibit some interesting and instructive peculiarities. The reduced estimate of the rate of depreciation results in net investment reaching zero at considerably older ages than with previous estimates. Because of the different rates of exogenous growth estimated for different schooling levels, the income patterns of those cohorts who were 20 years and 10 years of age in 1960 show lifetime incomes which do not increase monotonically with increasing schooling levels. We conclude that this must imply that the trends for middle-aged workers, which dominate the parameter estimates, cannot continue if

* Assistant professor of economics and statistics, Southern Methodist University, and econometric analyst, Weyerhauser Company, Tacoma, Washington, respectively. The research for this paper was supported by the Department of Labor Manpower Development Grant 31-46-70-06 to Southern Methodist University. We are grateful to Theodore W. Schultz, William J. Haley, George Borts, and an anonymous referee for helpful comments on an earlier draft. However, we alone are responsible for remaining deficiencies.

[1] See for example Gary Becker (1964, 1967), Giora Hanoch, Lee Hansen, Barry Chiswick, Becker and Chiswick, Johnson (1970), Jacob Mincer (1962).

[2] See Theodore Schultz (1961), Edward Denison, and the critique by Mary Jean Bowman. In the applications of the human capital model to the analysis of growth, Bowman has emphasized the distinction between internal rate of return and rental rates to human capital.

[3] See particularly Mincer (1970), and Mincer and Chiswick as well as the other papers in Schultz (1972).

[4] Mincer (1970, 1971) claims some superiority in the fit to his data when he uses the exponentially declining form of the *OJT* investment function as compared to the linearly declining form used by Johnson (1969, 1970).

monetary incentives to higher levels of schooling are maintained.

I. The Investment Model

Following Johnson (1970, p. 549), we write $E(\theta, \mathrm{t})$ for the earning capacity of an individual with given characteristics who is a member of cohort θ, and who is t years past the date at which the investment decision was made. Then

$$(1) \quad E(\theta, \mathrm{t}) = B \exp \left\{ - \int_0^\theta g(w)dw + \int_0^{\mathrm{t}} [R(u)k(u) - \delta(u)]du \right\}$$

$$Y(\theta, \mathrm{t}) = (1 - k(\mathrm{t}))E(\theta, \mathrm{t})$$

where

$Y(\theta, \mathrm{t})$ = observed earnings under the assumption that all investment costs are in the form of foregone earnings

B = the base earning capacity which the individual of cohort $\theta=0$ would have at the time the decision was made to invest in human capital $\mathrm{t}=0$

$g(\mathrm{t})$ = the rate of growth in earning capacity which would accrue irrespective of further investment or depreciation

$R(\mathrm{t})$ = the rental rate on all investments in human capital made at age t

$k(\mathrm{t})$ = the fraction of earning capacity which is invested in human capital made at age t

$\delta(\mathrm{t})$ = the rate of depreciation of the total stock of human capital possessed at t

The crucial assumptions in the derivation of the model in equation (1) are that the rental rate $R(\mathrm{t})$ on an investment at t remains constant on the undepreciated portion until retirement; that all human capital held at t depreciates at the same rate $\delta(\mathrm{t})$; that exogenous growth $g(\mathrm{t})$ accrues equally to the value of all human capital held at t; and that $\delta(\mathrm{t})$ and $g(\mathrm{t})$ are independent of each other and all other parameters in the model.

Johnson estimated the model in equation (1) using cross-section data for 1960 under the assumptions that $g(\mathrm{t})+\delta(\mathrm{t})=D$ = constant, $R(\mathrm{t})=R$ = constant, and $k(\mathrm{t}) =\alpha-\alpha(\mathrm{t}-\mathrm{S})/(\mathrm{T}-\mathrm{S})$ for $\mathrm{t}-\mathrm{S}>0$, where S = years of schooling and T = predetermined age at retirement with retirement assumed to occur at the sixty-fifth birthday.

Depreciation of the stock of human capital is entered in the model such that it does affect the pattern of an individual's (or cohort's) lifetime income. The exogenous growth parameter is entered such that it affects the magnitude, but not the pattern of an individual's lifetime income profile. When the parameters are estimated from a single cross-section of cohorts of different ages, the effects of depreciation and exogenous growth cannot be distinguished (see Johnson, 1970). If time-series data were available on the earnings of a single cohort throughout the earning lifetime, the exogenous growth term would enter only as an adjustment to the base earnings term B if $\theta=0$ corresponded to a different cohort. If such time-series data were available for two or more cohorts, an estimate of the exogenous growth parameter of equation (1) could be obtained from the changes in the base earnings between cohorts. In the absence of time-series over the entire earning lifetime of several cohorts, a series of cross-sectional data are used even though the time-series on any one cohort is much less than the span of the earning life. The results show that the exogenous growth parameters specified are identified empirically by these data.

Previous researchers have also recognized that earnings increases which are not the result of increased years of school-

ing or experience cause a cross-section of earnings data to give an upward bias to estimated returns to human capital. Becker (1964) has approached the problem by estimating the corrected rate of return to education for several different rates of exogenous growth. Thus, while it is not fair to accuse previous researchers of presenting only biased estimates of rates of return, we believe that the technique employed in this paper is an improvement over the approach required when only a single cross-section is used.

There are several possible sources of autonomous growth including increases in the ratio of nonhuman capital to human capital, or autonomous shifts in the production functions. Also, if there is less than perfect substitution between individuals in different cohorts, the relative scarcity of the 1930's cohort could result in increases in earnings to that cohort which would appear in our model as exogenous growth. The empirical results which we obtain are consistent with the hypothesis that increasing the average level of schooling produces social externalities which accrue more (in terms of percentage increase in earnings) to the less educated. Of course here again this effect may be due more to decreasing the supply of persons with little schooling than to increasing the demand for their services. However, we have not attempted a thorough analysis of the sources of exogenous growth and we avoid this question for the remainder of the paper.

It is well known that with a finite life expectancy the rental rate must vary with the age at which the investment is made if a constant internal rate of return is to be maintained. However, when the form $R(t)=r/[1-\exp(-t(T-t))]$, which maintains a constant internal rate of discount r, is substituted into (1), a much more complicated expression for earnings results. A term which is an infinite series is particularly troublesome in the computational procedure.[5] Johnson's (1970) results for which this correction should be most important were replicated and in none of these cases did the finite life corrections effect more than a 0.0007 change in the estimate for r from the corresponding estimate of R. We thus conclude that in the more extensive calculations estimating the growth rate, the added complication of finite life correction may safely be ignored. This also implies that the results reported previously by Johnson (1970) do not differ from other results because of the absence of a finite life correction. We now turn our attention to the error terms which must be added to equation (1).

II. The Statistical Model Combining a Time-Series of Cross-Sections of Data Error Components

With these types of data it is well recognized that several components of errors are likely to be present. Some special cases of interdependence of error components in time-series and cross-sectional data have been explored in the econometric literature.[6] However, none of these econometric papers develops a model in which serial correlation may be present in several dimensions. The disturbances associated with each cohort contribute to the deviation of each observed datum from the specified model. However, the background conditions which affect investment decisions do not change completely nor randomly from one cohort to the next, leading to serial correlation among the errors along a cross-section.[7]

[5] However, it is estimated that the four-term truncation of the series which was used yielded accuracy within 4 percent for the term which represented a small correction. See Hebein (1972).

[6] See, for example, Dudley Wallace and Ashiq Hussain, Marc Nerlove (1967, 1971a,b) and G. S. Maddala. Nerlove (1971a) analyzes a model which allows for serial dependence among the disturbances for the same individual, but no dependence for different individuals.

[7] Another source of potential serial correlation of er-

Similarly, the varying effects of aggregate economic conditions upon the observed earnings in different cross-cohort sets are not explained by the model. Since aggregate economic conditions change relatively slowly when cross-sections are spaced as closely as two-year intervals we expect serial correlation of errors along each cohort. And finally, similar disturbances affect men at the same points in their lives; for example, labor force entry, and family formation. These forces, unexplained explicitly by the model, lead us to anticipate serial correlation of errors when observing successive cohorts within a schooling group at some fixed birthday.

If a logarithmic transformation makes the errors additive,

$$Y^*_{t\theta} = \log Y_{t\theta} + U_t + V_\theta + W_{t\theta} \tag{2}$$

where

- t = the cross-section $t=1, \ldots, T$
- θ = the cohort $\theta=1, \ldots, N$
- $Y_{t\theta}$ = "observed earnings" from equation (1)
- U_t = the variance component associated with the t-th cross-section
- V_θ = the variance component associated with the θth cohort
- $W_{t\theta}$ = the variance component associated with the θth cohort in the t-th cross-section

Assuming that the serial correlation is of the first-order autocorrelation form and is confined to the W error component, we write

$$W_{t\theta} = \rho_a W_{t-1,\theta-1} + \rho_b W_{t-1,\theta} + \rho_c W_{t,\theta-1} + \gamma_{t\theta} \tag{3}$$

where the first term is associated with individuals of the same age, the second with the θth cohort, and the third with the t-th cross-section. If the $W_{t\theta}$ component dominates the error in equation (2) then the linear transformation

$$\gamma_{t\theta} - W_{t\theta} - \rho_a W_{t-1,\theta-1} - \rho_b W_{t-1,\theta} - \rho_c W_{t,\theta-1} \tag{4}$$

yields an approximately linear covariance matrix. Hebein has shown that under these assumptions E. J. Hannan's results assure consistency and asymptotic normality of the parameter estimates.

Statistical Model

The statistical model is obtained from equation (1) by taking the logarithmic transformation of $Y(t, \theta)$, assuming the linear form for $k(t)$ as did Johnson (1970) and letting $\delta(t)=\delta$, $R(t)=R$, $g(\theta)=g_0+2g_1\theta$.[8] Thus

$$\begin{aligned} Y^*(t, \theta) &= \log Y(t, \theta) \\ &= \log (B(1.0-\alpha+\alpha\tau/(T-S))) \\ &\quad +(R-\delta)S+(R\alpha-\delta)\tau \\ &\quad -R\alpha\tau^2/2(T-S)-g_0\theta-g_1\theta^2 \\ &\quad +W_{t\theta} \end{aligned} \tag{5}$$

where

t = current age in years since decision

rors in a cross-section is the manner in which the data taken from the *Current Population Reports* were developed. The Bureau of the Census estimates annual earnings for different schooling levels by fitting a parabolic function to mean incomes of the age intervals 18–24, 25–34, 35–44, 45–54, 55–64, and 65 years of age and older. Although we have used the published yearly earnings as if they were independent observations, it must be recognized that each "observation" is explicitly related to adjacent observations. One use of the model presented here might be to improve the interpolation procedures used by the Bureau of the Census. For a more complete and geometrically illustrated discussion of these error components and the potential serial correlation, see Hebein. Including terms beyond the constant in the function $g(\theta)$ allows for a greater (or lesser) exogenous growth rate to accrue to younger cohorts; a smaller θ corresponds to a younger cohort. However, exogenous growth still accrues once and for all to each cohort.

[8] The deviation of the model does not allow for exogenous growth to an individual's earnings to fluctuate over his lifetime. Such fluctuations would have to enter in the $R(t)$, $k(t)$, or $\delta(t)$ functions.

S = years of schooling after decision
$\tau = t - S$
T = retirement age
C = calendar date of cross-section, for example, 1956, 1958, . . .
θ = cohort defined by the t age in 1959 ($\theta = t - C + 1960$) the transformation defined by equation (4) is applied

In the following section we introduce the data used in estimation, present the estimates obtained for the parameters α, R, δ, g_0, and g_1, and develop some implications of the estimated values of these parameters.

III. Estimation

A. *Data*

Data on income by age and schooling level are contained in the *Current Population Reports* (*CPR*). Since the model is developed to represent observed earnings, use of total income data will bias the estimates of returns to schooling due to the well-established correlation between property income and level of schooling attained. Taking 1960 as the base year and using the base earnings capacity B, estimated previously by Johnson (1970), means that inclusion of nonearnings income will bias the estimated returns to schooling upward. However, the results indicate that this is not a severe problem with these data.

The *CPR* tabulate estimated annual income of males with income during the survey year for the ages 18 to 64 by schooling level completed for the years 1956, 1958, 1961, 1963, 1964, and 1966. The Census Bureau estimates each observation of income by fitting a parabola to mean income data of six age groups. The incomes predicted by the parabolic function are adjusted by mortality rates and price level and the resulting estimates of income are published in Tables 7–13 in the *CPR*. Thus, all income data are in 1958 dollars and the increase in the Consumer Price Index does not affect the estimate of the rate of exogenous growth.

Thus, there are three main problems with these data: Total income rather than earnings are reported; the parabolic fit; and, since the youngest age reported is 18 years, base earnings B must be obtained from other sources. The process of aggregation into six age groups and subsequent disaggregation into forty-seven years of income records means that we have at least six and at most forty-seven degrees of freedom in each. Although our model can be integrated over age intervals to utilize age group mean data directly, we have chosen to use the more readily available yearly income figures. While this choice introduces some inefficiencies into the parameter estimates reported here, it demonstrates the ability of our procedure to handle larger quantities of data which is always a concern with non-linear estimation procedures.

There are several additional qualifications to our empirical estimates which should be noted explicitly. First, no explicit corrections are made for ability nor for the average days of school attendance. The effects of increases over time in average days of school attendance will be reflected in our estimates of the "exogenous growth" parameters.

We also do not explicitly separate out "employment" effects. The conditions of the economy and of the labor markets in particular will affect income and hence the estimates. However, our goal is more modest than that of Schultz (1961) or Denison. We are analyzing the effects of investments in human capital upon personal earnings (income)[9] and thus avoid

[9] We are also accepting the biases resulting from longer hours of work of more educated persons (see C. M. Lindsay) and the different levels of nonmonetary benefits of employment which tend to counteract the longer work hours. We do not attempt here to answer

many of the problems of national income growth accounting.

B. *Estimation of Parameters*

The modified Hartley procedure (see Nelson et al.) is used to estimate the model parameters. Since the first-order autocorrelation coefficients, ρ_a, ρ_b, ρ_c, are not known a priori these coefficients are included as parameters to be estimated simultaneously with the parameters α, R, δ, g_0, and g_1 with the iterative procedure.

When ρ_c is included in the regression, the estimated value is very near to unity. Essentially, inclusion of ρ_c results in first differencing of the dependent variable, causing the *OJT* parameter α and the depreciation rate δ to become very small. Thus, the inclusion of ρ_c in the estimation essentially eliminates the influence of the analytical model leaving lagged income as the only remaining "explanatory" variable in the regression. Therefore, this parameter is not included in the following estimates.

C. *The Effect of the Gross Investment Function*

It would appear that the serial correlation of residuals, which could not be removed effectively with the use of a first-order serial correlation coefficient, might be removed with a minor modification of the estimated form of the model. Using the linear gross investment function, the regression predicts incomes that are too low early and late in life while overpredicting the income in middle years. The high correlation coefficients obtained using the linear form indicate that the linear gross investment function used does provide a good approximation to investment, leaving little room for improvement in the sum of squared errors criteria. However, it also appears that some improvement might be possible in the serial correlation along the cross-sections.

Several functional forms of the gross investment function have been proposed by other researchers. Mincer (1971) suggests that an exponential form of the gross investment function after the completion of schooling is superior to the linear, negatively sloped gross investment function. Similarly, Haley derives a form of the gross investment function which resembles a third degree polynomial. We tested two alternative forms of the postschool gross investment functions. A polynomial form was tried both with gross investment forced to zero at retirement (assumed to be at the sixty-fifth birthday) and with gross investment allowed to be positive at retirement. The exponential form was not forced to zero at retirement.[10]

With an *ad hoc* procedure of adding to the sum of squared errors a penalty function proportional to the deviation of the Durbin-Watson statistic from the value 2.0, we found that we could obtain *D-W* values of approximately two while reducing the R^2 from 0.9 to 0.74. Most parameter estimates were little affected, but the value of α declined from 0.373 to a constrained value of 0.271. Starting from the parameter estimates thus obtained, the model with the exponential gross investment function $k(t)$ was estimable. However, including the exponential form of $k(t)$ did not reduce the serial correlation

the question whether increased labor force participation and hours of work and later retirement are a cost or a benefit of increased human capital.

[10] Unfortunately, the estimation of neither the polynomial nor the exponential form was possible with the non-linear algorithm using the best linear estimates as initial values. The computer program constructs an information matrix with respect to the parameters by calculating the matrix of second-order partial derivatives. The inverse of the matrix must exist at each iteration if estimates of the parameters are to be obtained. Computation terminated prior to the completion of the first iteration through the mode of linear dependence in the information matrix.

TABLE 1—ESTIMATES OF MODEL PARAMETERS[a]

L	U	B	α	R	δ	G_0	G_1	ρ_a	ρ_b	DF[c]	DW	R^2
5–7	8	1750	.430	.275	.013	.053	−.00003	.756	.0018	275	.055	.950
			.0007	*.0001*	$<10^{-4}$[b]	$<10^{-4}$	$<10^{-8}$	*.0007*	*.0025*	(29)		
8	9–11	1900	.442	.288	.019	.053	−.433	.783	.331	275	.054	.962
			.0002	*.0001*	$<10^{-4}$	$<10^{-4}$	$<10^{-7}$	*.0007*	*.002*	(29)		
8	12	1900	.383	.304	.033	.046	−.0006	.133	.541	263	.055	.945
			.0001	$<10^{-4}$	$<10^{-5}$	$<10^{-5}$	$<10^{-8}$	*.002*	*.0007*	(29)		
9–11	12	2510	.311	.380	.032	.046	−.0006	.132	.543	263	.056	.945
			$<10^{-4}$	*.0001*	$<10^{-5}$	$<10^{-5}$	$<10^{-8}$	*.004*	*.0007*	(29)		
12	13–15	3920	.447	.259	.034	.044	−.0005	.415	.516	245	.111	.977
			$<10^{-4}$	10^{-5}	$<10^{-5}$	$<10^{-8}$	$<10^{-6}$	$<10^{-5}$	*.002*	(29)		
12	16	3920	.373	.177	.010	.029	−.0002	−.112	.040	188	.054	.899
			.0004	$<10^{-4}$	$<10^{-4}$	10^{-5}	$<10^{-8}$	*.003*	*.0025*	(23)		
13–15	16	7160	.349	.190	.010	.029	−.0002	−.113	.042	188	.055	.900
			.0004	*.0002*	$<10^{-4}$	10^{-5}	10^{-8}	*.009*	*.003*	(23)		

[a] L=Lower schooling level; U=Upper schooling level; B=Base earnings (from Johnson 1970); α=OJT parameter; R=rental rate; δ=depreciation rate; G_0=constant term in growth; G_1=coefficient of cohort age; ρ_a and ρ_b=autocorrelation coefficients; DW=Durbin-Watson statistic along cross-sections; R^2=coefficient of determination.

[b] Very small values of standard errors of estimates are given as upper bounds (see Hebein for the precise values).

[c] DF=Degrees of freedom. If only the groups are counted as observations, the degrees of freedom are those recorded in parentheses.

along the cross-section, nor did it increase R^2.

Consequently, in the reported regressions (which estimate autonomous growth and cross-sectional parameters for all schooling levels considered) the linear, negatively sloped gross investment function is used. For theoretical and computational reasons discussed earlier, only the serial correlation coefficients ρ_a and ρ_b are included in these regressions. Unfortunately, the serial correlation along cross-sections as measured by the Durbin-Watson statistic remained serious. The results of these regressions are recorded in Table 1 and are analyzed in the next section.

IV. Analysis of Results

In previous studies using cross-sectional data to estimate rates of return, the estimates were necessarily biased by the rate of autonomous growth in earnings unless some estimates of the growth rate were obtained exogenously. Johnson (1970) has shown that in the model used in this paper the net rate biases the estimate of the depreciation rate δ upward when only cross-sectional data are used. The magnitude of these effects can be seen from the parameter estimates presented in Table 1.

Several interesting relationships can be seen by comparing these results with those previously obtained with this model using a single cross-section of data. First, since both the depreciation rate δ and the constant rate in the growth rate g_0 are positive, the true net rate of return $R-\delta$ lies between the gross rate of return R and the biased net rate $R-\delta-g_0$. The last figure $R-\delta-g_0$ is more comparable to the rate of return computed by Hanoch.

The larger values of net return in the present study are due primarily to smaller

values of depreciation. The present estimates of depreciation do not include the effects of linear autonomous growth. Thus, the estimated depreciation rate is consistently smaller relative to that measured with only cross-section data. Except for the terminal schooling levels 12 and 13–15, the sum of depreciation and autonomous growth $(\delta+g_0)$ is smaller than the depreciation rate estimated by Johnson for whites in the North. The sum $\delta+g_0$ is about 25 percent of the rate of return R for high school graduates, and nearly 30 percent of the rate of return R for individuals with a few years of college. Thus, Johnson's previous estimates of net returns $(R-\delta-g_0)$, which did not isolate the effects of depreciation and growth, are biased downward. The net rate of return for 13–15 years of schooling which we have estimated is 29 percent higher than the net rate estimated by Johnson and nearly three times as large as the rate of return estimated by Hanoch. The net rate of return of 0.225 for 13–15 years of schooling compares favorably with the 0.167 and 0.271 rates of return for 16 and 12 years of schooling, respectively.

Depreciation δ is greatest for high school graduates and for those with a few years of college. This is of particular interest since these educational levels provide much of the skilled labor in the United States and are considered complementary to physical capital (see Finis Welch). The greater depreciation encountered by these individuals suggests that their skills may be more sensitive, relative to individuals at other schooling levels, to capital innovation.[11]

The values of the growth parameter g_0 decline with increasing schooling. The decreasing values of g_0 imply that schooling and *OJT* account for larger portions of the explained earnings as schooling levels increased.

Hypotheses on parameter values may be tested using the approximate F suggested by Johnson (1970). With this procedure the hypothesis that the linear growth parameter g_0 is equal for 8 and 12 years of schooling cannot be rejected $(F<1)$; however, the hypothesis that g_0 is equal for 12 and 16 years of schooling is rejected at the 1 percent level of significance. Thus schooling and *OJT* determine significantly different amounts of explained earnings for the schooling levels 12 and 16. The results of the tests of hypotheses indicate that beyond the twelfth year of schooling, schooling and *OJT* increase in importance in the explanation of personal earnings since autonomous growth declines.

The average proportionate increase in cohort earning profiles not explained by schooling or *OJT* is measured by autonomous growth, which collects additions to earnings from all other sources. While there are several possible sources of autonomous growth, any theory attempting to explain a portion of autonomous growth should, ideally, explain the decline of g_0 with increased schooling. Our model was not designed to include the macro-economic effects relating to growth and we will not here pursue possible explanations of this growth phenomenon.

Additional insights into the meaning and plausibility of the parameter estimates can be obtained by calculating the implied magnitudes of total lifetime income, total and net investment in schooling, and total and net investment in *OJT*. These calculated values are presented in Table 2 together with similar values calculated from

[11] The high rate of depreciation suggests that these individuals might maximize earnings by spending shorter, more frequent, periods in full-time training relative to individuals with lower depreciation rates. Alternatively, we might expect larger fractions of their earnings capacity to be invested (gross) in *OJT*. However, no pattern of consistently high investment in *OJT* is observed in the values of α, Table 1, although α is unusually large for the 13–15 year group.

TABLE 2—QUANTITIES DERIVED FROM PARAMETER ESTIMATES

Cohort Group	Lower Schooling Level	Upper Schooling Level	Lifetime Income	Gross Schooling Investment	Net Schooling Investment	Gross *OJT* Investment	Net *OJT* Investment	Birthday at which Net Investment=0[d]
Whites/North:	5–7	8	$ 199,000	$ 4,320	$ 2,960	$ 91,600	$ −30	40.5
1960 Census	8	9–11	220,000	4,910	3,760	68,300	900	42.1
Cross-Section[a,b]	8	12	245,000	12,040	9,410	67,700	−690	42.1
	12	13–15	278,000	15,470	11,290	93,900	−6,600	41.7
	12	16	342,000	38,900	27,800	115,700	−14,500	41.8
47 Years Old	5–7	8	173,000[c]	1,130	1,080	31,300	21,600	58.7
in 1960	8	9–11	208,000	1,540	1,440	40,000	23,900	57.2
	8	12	257,000	6,060	5,400	50,000	16,300	51.4
	12	13–15	295,000	7,380	6,590	70,400	22,700	52.1
	12	16	389,000	23,700	22,300	75,400	47,700	57.9
35 Years Old	5–7	8	281,000	1,840	1,750	50,900	35,100	
in 1960	8	9–11	310,000	2,300	2,150	59,700	35,650	
	8	12	332,000	7,820	6,970	64,500	21,000	
	12	13–15	412,000	10,600	9,200	98,200	31,700	
	12	16	517,000	31,500	29,700	100,100	63,400	
20 Years Old	5–7	8	571,000	3,740	3,560	103,300	71,300	
in 1960	8	9–11	611,000	4,530	4,240	117,500	70,200	
	8	12	582,000	13,700	12,200	113,100	36,800	
	12	13–15	750,000	19,300	16,800	179,000	57,800	
	12	16	793,000	48,400	45,500	153,800	97,400	
10 Years Old	5–7	8	975,000	6,380	6,080	176,200	121,700	
in 1960	8	9–11	1,070,000	7,940	7,420	205,700	122,800	
	8	12	983,000	23,100	20,600	190,900	62,200	
	12	13–15	1,253,000	32,200	28,000	299,000	96,600	
	12	16	1,105,000	67,500	63,400	214,100	135,700	

[a] Taken from Johnson's (1970) analysis of the 1960 Census cross-section.
[b] The 9–11/12 and 13–15/16 schooling groups have been omitted because these results add little to the implications.
[c] All income and investment numbers are in 1958 dollars.
[d] The birthday at which net investment is calculated to reach zero does not change with cohort.

the 1960 Census cross-section data by Johnson (1970). Comparing the 1960 Census results with the cohort values shows that total lifetime incomes for the cohort that was 47 years old in 1960 compare favorably with the Census cross-section. The 47-year old cohort is near the middle of the labor force lifetime and it is not surprising to find this agreement. Examination of the calculated gross and net investments in schooling reveals that the values calculated for the 20-year old cohort agrees more closely with the cross-section estimates. Again, this is not surprising since schooling investments are made immediately following the investment decisions and for each schooling level the cohort age θ was scaled to zero for those at the decision age in 1960. Thus, since the base earnings B used in both this analysis and the earlier cross-section study was inferred for those of decision age in 1960, the growth term leads to lower base earnings (and consequently lower foregone earnings as investment in schooling) for older cohorts. The gross *OJT* investment values calculated for the 35-year old cohort compare more favorably with the cross-

section estimates. This also seems reasonable since our model and estimates put the highest rates of investment in *OJT* immediately after leaving school resulting in the average age of *OJT* investment being less than the average working age but, of course, greater than the decision age. In the present estimates we find proportionately less difference in gross and net investments due to the lower depreciation rates estimated in the present study. This lower depreciation rate is the major cause of the increase, in comparison with the earlier analysis, in the age at which net investment equals zero.

Everything seems reasonable until we examine the lifetime income versus schooling profiles of those persons 20 and 10 years of age in 1960. It does not seem reasonable that the economy will develop such that high school dropouts will have higher lifetime incomes than high school graduates and college dropouts have higher lifetime incomes than college graduates. These calculated values result from the different values of exogenous growth estimated for the different schooling levels while the parameter estimates obtained are dominated by those cohorts of persons who had income for the entire 1956–66 period. Therefore it is really an heroic extrapolation to compute lifetime incomes for persons not receiving income during the 1956–66 period. However, these extrapolations are of interest since they indicate, by the anomalies mentioned above, that the estimated growth trends probably will not (have not) continue.

Furthermore, even though we deprecate the accuracy of the extrapolations of lifetime income, they give us explicit examples in cautioning against some common misuses of lifetime income estimates. For example, how accurate would have been the following statement in 1960 to a tenth grader contemplating dropping out of high school: "Finish your education; why just look, it will be worth $25,000 to you in added income over the rest of your life"; or a similar statement to a potential college dropout in 1970?[12]

V. Summary and Conclusions

In this paper we have made operational the suggestion of Johnson (1970) that with time-series observations the rate of growth in earnings could be estimated simultaneously with the rate of return and the quantity of on-the-job training. We found that in this model the correction for finite life makes no significant impact on parameter estimates; thus this rather burdensome correction is not included in the main empirical work. We attacked the problem of error components in the combination of time-series and cross-section data. However, we were not successful in reducing autocorrelation of residuals in cross-section even though we investigated several functional forms of the *OJT* investment function. In particular, the exponential form of the investment function yielded no improvement in either R^2 or the Durbin-Watson statistic measured in cross-section.

Using the *Current Population Reports* data of personal income for 1956–66 the estimates of rates of return, depreciation, and *OJT* investment are in reasonable agreement with previous estimates. With removal of the bias of exogenous growth, which is present in Johnson's cross-section estimates, the estimated rates of depreciation are more reassuring than the very

[12] Our faith in the orderly working of markets and knowledge that returns to high school and college levels of schooling have remained attractive over several decades would give us confidence in arguing that these would be good investments, although we would now be most hesitant in quoting any fixed monetary sum (especially if we had obtained that sum from cross-sectional data alone). In the estimation procedure it is clear that we have assumed that gross rates of return to each level of schooling R remains constant over the years with changes showing up in the growth paramameters.

high rates previously estimated by Johnson. When some of the implications of the parameter estimates are calculated, this reduced rate of depreciation results in net investment reaching zero at considerably older ages than with previous cross-section estimates.

The lifetime incomes by cohort implied by the parameter estimates exhibit some interesting and instructive peculiarities. Because of the different rates of exogenous growth estimated for different schooling levels, the income patterns of those cohorts who were 20 years and 10 years of age in 1960 show lifetime incomes which do not increase monotonically with increased schooling levels. We conclude that this must imply that the trends for middle-aged workers which dominate the parameter estimates cannot continue if monetary incentives to higher levels of schooling are maintained.

Finally, we believe that further work on the combination of time-series and cross-sectional data would be desirable with emphasis on serial correlation in several dimensions. In conjunction with this work, a multidimension Durbin-Watson type statistic would be very useful.

REFERENCES

G. Becker, *Human Capital*, New York 1964.

———, "Human Capital and the Personal Distribution of Income," W. S. Woytinsky lecture no. 1, Inst. Public Admin., Univ. Michigan 1967.

——— **and B. R. Chiswick**, "Education and the Distribution of Earnings," *Amer. Econ. Rev. Proc.*, May 1966, *56*, 358–69.

M. J. Bowman, "Schultz, Denison, and the Contribution of 'Eds' to National Income Growth," *J. Polit. Econ.* Oct. 1964, *72*, 450–65.

B. R. Chiswick, "Human Capital and the Distribution of Personal Income," unpublished doctoral dissertation, Columbia Univ. 1967.

E. F. Denison, "Education, Economic Growth, and Gaps in Information," *J. Polit. Econ.*, Oct. 1962, Suppl., *70*, 124–28.

W. J. Haley, "Human Capital Accumulation over the Life Cycle," unpublished doctoral dissertation, No. Carolina State Univ. 1971.

E. J. Hannan, "Non-Linear Times Series Regression," *J. Applied Probability*, Dec. 1971, *8*, 767–80.

G. Hanoch, "Personal Earnings and Investment in Schooling," unpublished doctoral dissertation, Univ. Chicago 1965.

W. L. Hansen, "Total and Private Rates of Return to Investments in Schooling," *J. Polit. Econ.* Apr. 1963, *71*, 128–40.

F. J. Hebein, "Investment in Human Capital: Growth of Earnings and Rate of Return," unpublished doctoral dissertation, Southern Methodist Univ. 1972.

T. Johnson, "Returns from Investment in Schooling and On-the-Job Training," unpublished doctoral dissertation, No. Carolina State Univ. 1969.

———, "Returns from Investment in Human Capital," *Amer. Econ. Rev.* Sept. 1970, *60*, 546–60.

C. M. Lindsay, "Measuring Human Capital Returns," *J. Polit. Econ.*, Nov./Dec. 1971, *79*, 1195–1215.

G. S. Maddala, "The Use of a Variance Component Model in Pooling Cross-Section and Time Series Data," *Econometrica*, Mar. 1971, *39*, 341–58.

J. Mincer, "On-the-Job Training Costs, Returns and Some Implications," *J. Polit. Econ.*, Oct. 1962, Suppl., *70*, 50–59.

———, "The Distribution of Labor Incomes: A Survey," *J. Econ. Lit.*, Mar. 1970, *8*, 1–26.

———, "Education, Experience, and the Distribution of Earnings and Employment," Nat. Bur. Econ. Res., New York 1971.

——— **and B. R. Chiswick**, "Time-Series Changes in Personal Income Inequality in the United States from 1939, with Projections to 1985," *J. Polit. Econ.*, May/June 1972, Suppl., *80*, 534–66.

A. C. Nelson, Jr. et al., *Evaluation of Computer Programs for System Performance Effectiveness*, Research Triangle Inst., N.C. 1967.

M. Nerlove, "Experimental Evidence on the

Estimation of Dynamic Economic Relations from a Time-Series of Cross-Sections," *Econ. Stud. Quart.*, 1967, *18*, 42–74.

———, (1971a) "Further Evidence on the Estimation of Dynamic Economic Relations from a Time-Series of Cross-Sections," *Econometrica*, Mar. 1971, *39*, 457–82.

———, (1971b) "A Note on Error Components Models," *Econometrica*, Mar. 1971, *39*, 383–96.

T. W. Schultz, "Investment in Human Capital," *Amer. Econ. Rev.*, Mar. 1961, *51*, 1–17.

———, "Human Capital: Policy Issues and Research Opportunities," in his *Human Resources*, Vol. VI, New York 1972.

T. D. Wallace and A. Hussain, "The Use of Error Components Models in Combining Cross-Section with Time-Series Data," *Econometrica*, June 1969, *37*, 55–72.

F. Welch, "Linear Synthesis of Skill Distributions," *J. Hum. Resources*, Summer 1969, *4*, 311–27.

U.S. Bureau of the Census, *Current Population Reports*, Series P-60, *Consumer Income*, No. 56, Washington 1968.

Optimal International Borrowing and Lending

By JAMES A. HANSON*

Recent years have witnessed a renewed interest in the question of the optimal level of foreign borrowing and lending, a problem which has intrigued economists since Adam Smith.[1] However, most of the recent treatments have been somewhat simplified in their assumptions about the determinants of international lending and highly mathematical in their approach. The intent of this paper is to redress the balance slightly by including a simple measure of expropriation risk in the determinants of international lending and by providing some economic explanations for the results obtained in both this paper and some of the other recent work.

Most work on international borrowing assumes that its levels are sensitive only to interest rates,[2] while the effects on the lender of the possible risks of national default and expropriations are usually ignored. However, these risks seem particularly important in the emotion-ridden case of direct foreign investment,[3] and in the government to government or quasi-government to government loans, so prevalent in the twentieth century. They also appear in the case of private portfolio loans to governments and quasi-government agencies which characterized nineteenth century capital flows.[4] Whether the loan takes the form of portfolio or direct investment, this risk of national default cannot be ignored by any international lender since the government of the borrower's country can always intervene in the outflow of national debt service payments and remittances, independent of a particular loan's profitability. The possibility of such an intervention by a government, responsible to only one of the contracting parties, is one of the major differences between international and interregional economic transactions.

This paper argues that one measure of the national expropriation or default risk is the ratio of foreign to nationally owned capital. From the point of view of both visibility and the relative difficulty of meeting debt service payments and profit remittances out of domestic product, this seems a reasonable measure. If risk does, in fact, depend upon this ratio, then an increase in the domestically owned capital stock would reduce risk and tend to stimulate more foreign investment. Although accumulation would also lower the interest rate and tend to reduce foreign investment, the risk-reducing effect could dominate. In that case domestic accumulation would, on balance, stimulate foreign investment. The recognition of the importance of risk also implies, contrary to the usual assumption, that even a very small country would be unable to increase its foreign borrowings without either paying higher interest rates, increasing its domestically owned capital stock, or both.

* Brown University. I would like to thank my former colleague, Robert Barro, for helpful suggestions without implicating him in the resulting product.

[1] See Adam Smith, Book II, ch. 5, and Keynes. More recently, P. K. Bardhan, Koichi Hamada (1966, 1969), Philip Neher and the author, and Murray Kemp (1964, 1969) have all treated the problem.

[2] One interesting exception is Bardhan's work which incorporates a rising marginal disutility of foreign capital in a borrowing country for nationalistic reasons.

[3] See Marvin Bernstein, Charles Kindleberger, and Raymond Vernon.

[4] See Ragnar Nurkse.

In Section I of this paper a model incorporating these possibilities is developed, and an historical sequence or stage theory of accumulation, borrowing, and lending is outlined. Section II is devoted to demonstrating that a unique long-run growth equilibrium may be reached even in this more general model of international lending. The sufficient conditions for this equilibrium are then shown to have a well-founded economic rationale.

Section III turns to the optimal rate of accumulation. It differs from other, more recent treatments in its assumptions about lending behavior, as well as in its assumption that national and foreign owners of capital receive the same return on their capital. In contrast to this paper, other recent treatments for purposes of mathematical simplicity have assumed that lending is determined by the interest rate alone.

While the discussion of optimality in this paper is confined to steady rate growth paths for each analysis, the results are the limiting cases or "turnpikes" of the optimality conditions obtained from more mathematically complex models. Moreover, these results are appealing to our economic intuition. In the context of the more general model of this paper, it is assumed that foreign lenders respond positively to rises in interest rates and to the ratio of national to foreign owned capital. It can be shown that a borrowing country should allow national accumulation to proceed somewhat further than the golden rule suggested by other commentators on the optimal level of borrowing. However, if the response to a rise in the ratio of national to foreign owned capital is negative, accumulation should be constrained to a less than golden rule level. Finally, a lending country should confine its accumulation below the golden rule level if it is large enough to affect its borrowing rates.

Section IV extends the model to the case discussed by most recent authors in which the government can affect both the capital-labor ratio and the rate of return on foreign-owned capital. Again confining the choice to golden ages and neglecting the problem of moving to the optimum, some optimality rules are obtained which are similar to those found elsewhere in the literature,[5] although they must be modified to take into account the risk variable. It is shown that all these rules are basically the result of simple cost-benefit calculations, i.e., that domestic accumulation should continue until the additional per capita production just equals the cost in foregone consumption. This equality should be attained while holding foreign loans or borrowing constant, through changes in the taxation of remittances, at the level at which the marginal returns on domestic capital and the marginal cost of foreign borrowing or the marginal return on foreign lending are equal. Of course, this implies the country should take advantage of any monopoly position by imposing a tax equating marginal costs and returns. Since the earlier argument regarding risk implies that even a small country will face an upward sloping supply curve of foreign investment (domestically owned capital held constant), this argument has much broader applicability than is usually supposed. Finally, the above arguments regarding the relation between costs and saving and the relation between benefits and marginal products represent a correction of the fallacy, first advanced by Adam Smith, that when returns are the same, capital located in the home country is preferable to foreign lending because of the former's effects on home country wages.[6] Rather, capital should simply be invested where the net value of its marginal product is greatest.

[5] For example, see Hamada (1969).

[6] See Smith, Book II, ch. 5.

I. Capital, Equity, and Debt

Consider a neoclassical economy producing only one good called domestic product. It is assumed that there are only two factor inputs, capital and labor, and that the production function is subject to constant returns to scale and diminishing returns to the variable factor. Expressed algebraically:

$$y = f[k] \tag{1}$$

where y = domestic product per laborer in the home country; k = capital per laborer located in the home country; $f' \geq 0$; $f'' < 0$.[7]

The nationality of the owners of the capital stock as well as its location is important. To identify the owners clearly throughout the paper, the terms equity and debt are used. Equity is defined as the net capital owned by the citizens of the home country, whether located at home or abroad. Following the usual terminology, if equity exceeds the local capital stock then the country is a net lender, while if equity is less than the home country capital stock then the country is a net debtor. These definitions naturally lead to the definition of debt as the domestic capital stock owned by foreigners, whether their claims take the form of portfolio loans or direct investment. Under the same definitions, if none of the home country capital (net) is foreign owned, the country is a lender and debt will be negative. These definitions can be expressed algebraically as:

$$k = e + d \tag{2}$$

where e = equity per home country laborer; d = debt per home country laborer.

The distinction between capital stock owned by citizens and by foreigners corresponds to the national accounting distinction between domestic product and national product. Domestic product refers to the value of output produced in the home country while national product refers only to the production by nationally owned factors of production. In the model developed here, it is assumed that domestic product is distributed through competitive factor markets, with each factor receiving its marginal product. Assuming no government interference with factor payments,[8] equity and debt will receive the same rate of return. Therefore, total rentals on capital are divided between nationals and foreigners in proportion to their relative ownership of capital. Thus, with no labor migration, the relation between domestic product and national product is expressed algebraically by

$$z = y - f'd \tag{3}$$

and by Euler's Rule

$$z = f' \cdot (W + e) \tag{4}$$

where z = national product per laborer; $W[k]$ = wage rentals ratio.

Turning to the determinants of borrowing and lending, it is assumed that one basic factor is the rate of return on capital. It is further assumed that as the home country increases its per capita creditor position,[9] the overseas rate of return

[7] Lowercase letters refer to per capita magnitudes. All functions are shown with brackets while parentheses are used for multiplications. Except where it would be unclear, differentiation is indicated by a prime.

[8] In Section IV this assumption is modified to permit government tax policy to drive a wedge between foreign and domestic rates of return.

[9] The international interest rate is assumed to be a function of per capita magnitudes in order to make the problem an interesting one. Assuming the rate to be a function of the absolute level of lending or borrowing would make no sense, as it is equivalent to assuming a static "rest of the world" which would eventually be overwhelmed by home country growth. Assuming diminishing interest rates for increasing per capita lending, or rising rates for increased per capita borrowing, means the world and the home country's population are growing at the same rate. Neither the rest of the world nor the home country becomes completely dominant and interest rates do not change unless per capita variables change.

(eventually) falls. On the other hand, as a debtor, the home country will be assumed to face a nonnegatively sloped supply curve of debt per laborer, *ceteris paribus*. However, to a foreigner investing in the home country, all else may not seem to be constant when the amount of debt relative to the total capital stock increases. Any loan from the citizens of one country to the citizens or government of another involves the risk of national default and expropriation, as well as the usual individual loan risk; it is this risk which provides the distinction between international and interregional lending. As debt increases relative to the capital stock, the "visibility" of foreign "dominance" increases, and it becomes more and more tempting for the home government to attack the foreign capitalists as "exploiters" of the country and expropriate their claims. Perhaps more importantly, as the debt to capital ratio increases, more and more resources which are potentially available for domestic consumption must be allocated to meet debt service payments.[10] The growing costs of meeting these payments provides greater and greater temptation to expropriate or default, particularly since the international lender has little recourse to the legal system and the costs of expropriation in terms of higher interest rates on future borrowings are probably independent of the size of the expropriation.

The assumed reaction of foreign investors to expropriation risk as measured by the ratio of debt to total capital stock should not surprise those who have followed the World Bank and AID discussions of the debt servicing capacity of *LDC*s.[11] Such discussions have certainly been watched by the multinational corporations, who continually worry about investment "climate" and exchange rate policy. Once the possibility of this reaction is admitted, some interesting results immediately follow. First, the assumption provides an alternative rationale for Harry Johnson's argument that capital inflows depend on the level of income as well as the rate of interest.[12] Second, the assumption explains why even small countries might face a rising supply curve of international capital, since attempts at additional borrowing would disturb the debt-capital ratio and increase risk. Finally, it must be recognized that increases in domestic equity by increasing the ratio of capital to debt and of national product to domestic product might decrease risk sufficiently to outweigh the negative interest rate effects of accumulation, thereby encouraging more foreign loans.[13] Thus, an

[10] We make the usual neoclassical assumption, expressed for example by Gerald Meier, that resources can always be costlessly allocated to obtain the necessary foreign exchange for debt servicing and assume no terms of trade difficulties. If either of these assumptions is violated, then, as Meier points out, the costs of making debt service payments rise. Therefore, increases in the debt-capital ratio would correspond to even greater risks of expropriation.

[11] One of the earliest statements of this view and the importance of the income to foreign capital ratio can be found in Gerald Alter. Later treatments are found in Dragoslav Avramovic and Ravi Gulhati, and in Avramovic and Associates and Gulhati. Just Faaland's discussion of Gulhati suggests that economic growth will probably provide the flexibility and responsiveness to policy manipulation which is necessary to mobilize domestic saving for debt servicing.

[12] Other authors have made some attempt to incorporate the risk factor, although in rather discontinuous fashion. For example, Hamada (1969) assumes that foreign borrowing cannot exceed a fixed fraction of income. This assumption, which is mathematically useful, could be interpreted as either nationalistic feelings or a discrete unwillingness by foreigners to own more than a fixed fraction of the domestic capital stock. In an earlier paper, Hamada (1964) assumed that the risk of foreign loans is invariant up to some level of borrowing and then becomes infinite.

[13] Of course, this scenario may not be completely true. A variety of situations could be imagined in which national accumulation increased the risk on debt. For example, suppose that increases in equity are accompanied by the development of a local capitalist class. In some cases the interests of foreign and domestic capitalists would coincide and the development of a group of politically powerful domestic investors would reinforce the risk-decreasing aspects of domestic ac-

increase in the borrowing country's domestic equity may not be a substitute for foreign debt, as earlier, simpler models suggest.[14]

Moving from the motivation for capital flows to their size, it is obvious that the concept of the equilibrium per capita holdings of debt also requires some clarification and extension. In closed neoclassical growth models it is usually demonstrated that there is a unique, equilibrium relationship between each stock of capital per laborer, all relative prices, and per capita outputs which clears all markets. Then it is assumed that one of these short-run equilibria prevails at every moment in time,[15] with domestic capitalists holding the existing stock of capital. Extensions of the closed neoclassical model to include international lending must necessarily involve an extension of this concept of short-run equilibrium to include foreign as well as domestic owners of the capital stock.

In order to make this extension, assume as discussed above that the net stock supply of external funds, d, is dependent on the debt ratio, d/k, and the marginal productivity of capital, f':

$$d = D[d/k, f'] \tag{5}$$

with $\quad f' = f'[k]$

For simplicity, the equity-labor ratio, e, can be treated as determined by past saving, and therefore, as a parameter in the short run:

$$e = \bar{e} \tag{6}$$

Since the capital-labor ratio is equal to the sum of debt and equity (equation (2)) we have the equilibrium condition

$$d = D[(k - \bar{e})/k, f'] = k - \bar{e} \tag{7}$$

or $\quad k = D[(k - \bar{e})/k, f'] + \bar{e}$

The set of equilibrium pairs of capital-labor and equity-labor ratios which satisfy equation (7) are graphed in Figure 1B in the k, e plane, while the corresponding equilibrium interest rate is shown as a function of k in Figure 1A.[16] Equilibrium is used here to refer to a situation in which all markets are cleared and foreign investors are just satisfied with their holding of debt at the given equity-labor ratios and interest rate, a concept analogous to equilibrium in the usual closed neoclassical models.[17]

As a graphical example of an equilibrium, Figure 1A shows that if OA were the equity-labor ratio, then Ok_1 would be

cumulation and encourage further international lending. For example, domestic capitalists would probably be against default and expropriation of foreign debt, feeling they might be the next to fall. In this case an increase in equity would further decrease the risk involved in holding debt. On the other hand, to improve their own position, domestic capitalists might be in favor of restrictions on foreign capital which would diminish the profitability of individual loans to the home country. If this attitude dominates the risk effect described above, an increase in domestic equity would tend to reduce the equilibrium level of debt which foreigners wish to hold at constant interest rates. However, in this case the effect of domestic accumulation simply reinforces the interest effect (negative). Since this reinforcement will have easily interpreted effects, the remainder of the paper concentrates on the case in which risk is reduced, leaving the other case to footnotes and the reader.

[14] See the paper by Neher and the author. Hamada (1964, 1969) assumes that international capital flows receive an interest rate which is different from the rate of return on capital located in the home country and subject to manipulation through government taxation. Therefore, strictly speaking, debt and equity are independent. However, Hamada also assumes that the rate of interest and level of debt are inversely related, which would make debt and equity substitutes under the assumption that debt and equity are paid the same interest rate.

[15] See Hirofumi Uzawa.

[16] Per capita values are used because at each moment in time there is a fixed population which must be fully employed and which, together with capital and equity, determine per capita levels. Of course, this relationship between debt and equity does not mean that either debt or equity is the independent variable, only that there is a jointly determined equilibrium relation between them, similar, for example, to the Hicksian IS equilibrium relation between investment and saving.

[17] Applied to a lending country this assumption means that saving permits the acquisition of the amount of debt necessary to attain the equilibrium.

the corresponding equilibrium capital-labor ratio. The distance Oe_1, determined by the vertical projection from the point x, which lies on the 45° line, measures per capita equity on the horizontal axis. Since per capita debt equals k_1-e_1 and the rate of return on capital is $Of'(k_1)$, in Figure 1A, the rectangle wvk_1e_1 represents debt service payments. Per capita domestic product is equal to the rectangle $Ok_1uf(k_1)/k_1$. Per capita national product is that area less the rectangle wvk_1e_1.

The shape of the equilibrium relationship between capital, equity, and debt is an interesting one. It seems likely that even if the home country had no equity, foreigners would provide some capital stock. Thus $k|_{e=0}=D[1, f'[k]]>0$. However, the sign of the total derivative of the equilibrium level of debt per capita with respect to equity is unclear. Equity accumulation tends to depress the interest rate, thereby driving away foreign capital. However, it also decreases the debt-capital ratio, thereby reducing risk and stimulating foreign ownership of domestically

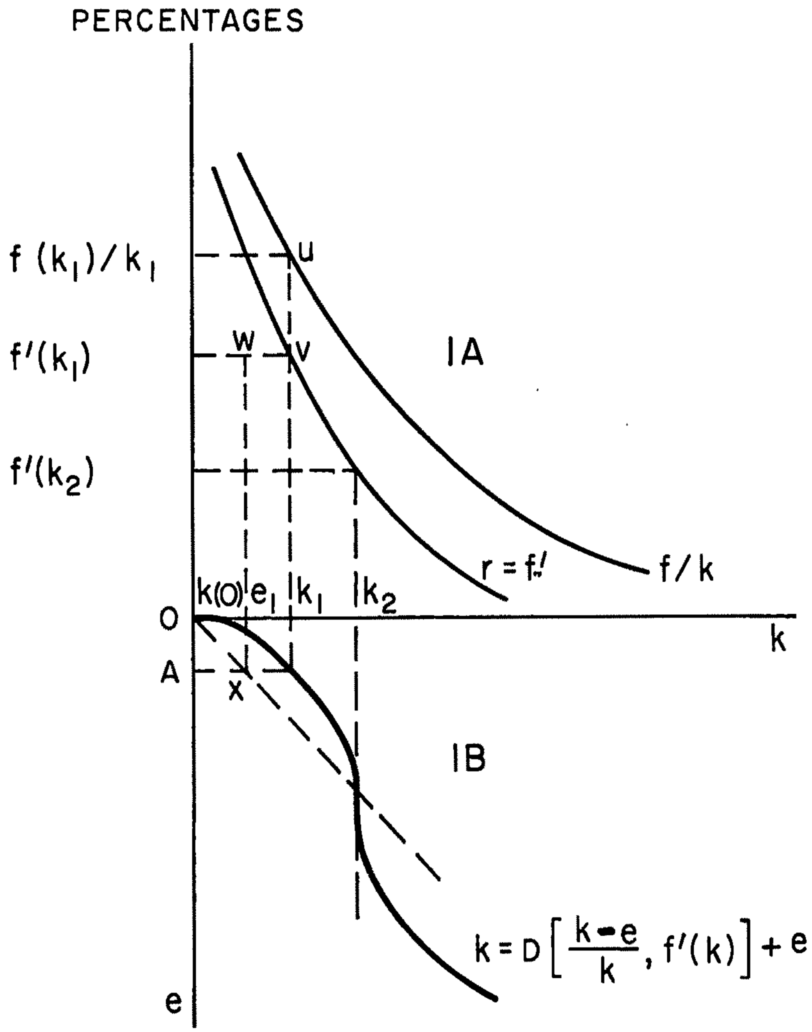


FIGURE 1

located assets. Assuming that the reduction in risk tends to dominate initially, but not by enough to raise the equilibrium debt-capital ratio at the new lower interest rate, we have

$$-1 \leq dD/de = D'$$

and $$0 \leq dk/de = k'$$

as the sign of the slope of the equilibrium locus, with

$$\epsilon_{ke} > \epsilon_{De}$$

where $$\epsilon_{ke} = k'e/k$$

and $$\epsilon_{De} = D'e/D$$

If we assume that a rise in the equity of a borrower does not lead to an equiproportionate rise in debt, then[18]

$$1 > \epsilon_{ke} > \epsilon_{De} \text{ with } d \geq 0$$

Based on these assumptions and following the previous discussion of the interaction between equity and debt, a typical history of growing per capita savings, equity, and debt accumulation might proceed as follows: at low levels of per capita equity, accumulation brings forth more, rather than less, per capita debt ($D' \geq 0$, $k' > 1$, $0 < \epsilon_{ke} < 1$) in spite of the fall in domestic interest rates. However, for foreign investors the favorable effects on risk of home country accumulation and a growing capitalist class are limited. Eventually the (negative) interest effects of equity accumulation begin to dominate, and debt and equity become imperfect substitutes ($-1 < D' < 0$, $0 < k' < 1$, $0 < \epsilon_{ke} < 1$). Finally they become perfect substitutes, with each increase in equity reducing the country's international debtor position or adding to the country's international creditor position in a one-to-one fashion, leaving interest rates unchanged ($-1 = D'$, $k' = 0$, $\epsilon_{ke} = 0$).[19] If accumulation continues, the country begins to affect international interest rates, although it is easy to imagine that at first there are sufficient markets for its loans to permit it to ignore their relative importance in the rest of the world ($-1 < D' < 0$, $0 < k' < 1$, $0 < \epsilon_{ke}$).[20]

The relationships described above are shown graphically in Figure 1. The equilibrium relationship $k = D + e$ begins at a positive capital-labor ratio ($k|_{e=0} = D[1, f'[k]]$. Initially equity accumulation stimulates an increase in debt, as shown by the growing horizontal distance between the 45° line and the equilibrium relationship, $k = D[(k-e)/k, f'] + e$. However, ϵ_{ke} is less than one, as graphically $e/k > k'$ when $d \geq 0$. Eventually the gap between the 45° line and the equilibrium locus narrows as debt and equity become substitutes. Finally they become perfect substitutes with the home country a small net borrower or lender, as shown by the vertical portion of the equilibrium locus, i.e., $k' = \epsilon_{ke} = 0$. The interest rate at which this occurs is shown in Figure 1A as $f'(k_2)$ and the intersection of the 45° line and the equilibrium locus indicates the equity-labor ratio at which the country is neither a net borrower nor lender. Of course, at larger equity-labor ratios, and lower domestic interest rates, the country begins to lend abroad.

[18] If equity accumulation increased the expropriation risk it might be possible for $k' < 0$ and the elasticity of capital with respect to equity, ϵ_{ke}, to be negative.

[19] This stage was discussed in my earlier paper with Neher.

[20] Strictly speaking, a completely symmetric treatment of capital movement would require a much more complicated model at a minimum, specifically treating the rest of the world by including the relative rate of diminishing returns to capital at home and abroad. By symmetry, when a home country becomes a lender, its investors might become concerned with the relationship between debt and the capital in the rest of the world, requiring foreign interest rates to exceed the domestic rate of return on capital to compensate for the risk. Though interesting, this point, as well as the relative rate of diminishing returns, is beyond the basically simple model of one country facing a world capital market contained in this paper. However, it can be noted that such risk-averse behavior would simply increase D' and ϵ_{ke}, but not change the limits.

II. The Accumulation of Equity

Equity increases through home country saving, assumed for simplicity to be a constant fraction, s, of national income,[21] while the effective labor force grows at the rate g. Thus

$$(8) \quad \frac{de}{dt} = sz - ge = sf'\cdot(W + k - d) - ge$$

where s = the saving rate.

Certain restrictions must be placed on the growth of the equity-labor ratio, described by equation (8), to ensure that the ratio neither degenerates to zero, nor rises without limit.[22] To prevent the first problem, the growth of equity per capita is assumed to be positive at very low equity-labor ratios; to prevent the second, the growth is assumed negative at very high equity-labor ratios. These restrictions also guarantee the existence of a stable steady state in which the growth of equity per capita is zero at intermediate equity-labor ratios; where equation (8) switches from positive to negative values. Mathematically the restrictions are:

$$\lim_{e\to 0} sf'\cdot\left(\frac{W}{e} + 1\right) = \lim_{e\to 0} sf'\cdot(\infty)\Big|_{d(0)=k(0)} + sf'\Big|_{d(0)=k(0)} > g$$

$$\lim_{e\to\infty} sf'\cdot\left(\frac{W+e}{e}\right) < \lim_{k\to\infty} sf'\cdot\left(\frac{W+k}{k}\right) = \lim_{k\to\infty} sf' < g$$

(assuming continuity and the relationship between k and e described above). These conditions are rather straightforward and somewhat less restrictive than in the closed economy case.

While the above conditions guarantee the existence of a stable long-run equilibrium, they are no assurance of its uniqueness. In particular, if equation (8) is not strictly decreasing in e, then multiple equilibria, some of which are unstable, may result. Such a problem may occur in the model of borrowing described here. To see this, assume the above limit conditions are met, so that at least one equilibrium exists. Differentiating equation (8) with respect to e, and substituting for g in the neighborhood of equilibrium, i.e., when $de/dt = 0$,

$$(9) \quad e\{f''\cdot(W+k-d)+f'\cdot(1+dW/dk)\}\cdot k' - f'D'e - f'\cdot(W+k-d) < 0$$

in the neighborhood of equilibrium to ensure stability. Noting $W = (f - f'k)/f'$, $W' = -f''f'(W+k)/f'^2$, $D' = k' - 1$, and finally dividing by k, we obtain:

$$(10) \quad \frac{-f}{k} + f' - f''\cdot\epsilon_{ke}\cdot d < 0$$

where $\epsilon_{ke} = k'e/k$.

Using the definition of the elasticity of substitution, σ, equation (10) yields

$$(11) \quad -\sigma(f'\cdot(W + k)/f'k) + \epsilon_{ke}\cdot d/k < 0$$

to ensure a unique, stable equilibrium.

It is obvious that if the country cannot affect world interest rates, and equity and debt exactly substitute for one another, then $\epsilon_{ke} = 0$ as was the case with earlier models. The first term of equation (11) is strictly negative and the equilibrium is unique and stable. In fact, this will be true whenever the net risk and interest effects of accumulation deter foreign loans, i.e., $\epsilon_{ke} < 0$.

If the (positive) risk effect of accumulation dominates the (negative) interest effect, as this paper assumes, there remain two possible conditions for a unique, stable, long-run equilibrium. On the one hand, if the saving rate is large enough to ensure that per capita equity growth is

[21] A saving rate dependent on the equity-labor ratio, as would be the case if capitalists saved more than laborers, could be analyzed easily.

[22] See Ken-Ichi Inada.

strictly positive while the country is a borrower, then a unique equilibrium as a lender ($d<0$) will be reached. Alternatively, if the elasticity of substitution exceeds the share of capital, then the equilibrium is also unique and stable.[23]

This last result also provides a clue to an intuitive explanation for the possibility of multiple equilibria. Neglecting the range in which borrowing does not affect world interest rates, the growth of equity raises the wage rate and lowers the interest rate. In a closed economy, the net result would be a fall in the average product of capital, but in an open, borrowing economy, the foreigner bears some of the decline in interest rates, yet does not share in the wage gains. It should be obvious that although the average product of capital still falls, the average national product and national saving per unit of equity may rise for a time, permitting the growth rate of equity accumulation based on saving out of national product to increase. This result is most likely to occur when foreigners hold a large fraction of the capital stock and maintain relatively constant holdings, when the share of capital is very large, and when the elasticity of substitution is very small. All these factors are incorporated in the sufficiency condition for uniqueness (equation (11)).

III. Optimal Accumulation

In earlier papers, owing to the simplified assumptions regarding capital flows and interest repayments, the optimal solution associated with the maximum (discounted) consumption per capita was similar to that reached in a closed economy, namely equality between the rate of interest and the population growth rates.[24] If necessary, borrowings were to be made up to some maximum allowable capital-debt ratio. The country would be a borrower or lender depending on the sign of $g-f'|_{k=e}$.[25] The assumptions of the present paper regarding factor payments lead to somewhat different answers.

For the sake of mathematical simplicity the discussion of optimality will be confined to a consideration of the choice between different proportional growth paths or golden ages, and neglect the obvious problems involved in deciding the rate at which the optimum should be approached.[26] Thus the results of this paper can be interpreted as the "turnpikes" of the more sophisticated models which involve no discount factor. Also, for the sake of simplicity it is assumed that each saving rate yields a unique long-run capital-labor and equity-labor ratio. The objective is the maximization of the sustainable level of per capita consumption,

$$(1-s)\cdot(f')\cdot(W+k-d) \tag{12}$$

subject to $de/dt=sf'\cdot(W+k-d)-ge=0$.

Substituting, differentiating with respect to e to obtain a maximum, and noting the definition of W', we obtain

$$(f')\cdot(k'-D')-f''k'd=g \tag{13}$$

[23] Even when equation (8) is not strictly decreasing in e, the long-run equilibrium between the growth rate of equity and population may be unique. For example, suppose that the production function is CES, $\sigma<1$ and the above condition does not hold at low equity labor ratios. Assume also that the saving rate is such that the growth rate of per capita equity is positive at low levels of per capita equity. However, as equity accumulation occurs, the first term of the expression is increasing in absolute terms and the second declining. Thus some maximum growth of per capita equity occurs at some positive level of equity growth. Beyond this level of per capita equity, the growth rate of accumulation is a strictly decreasing function of e. Therefore, the long-run equilibrium is unique.

[24] We neglect the rate of time preference used, for example, in Hamada (1969).

[25] See the papers by Neher and the author, and Hamada (1964, 1969).

[26] See Ivor Pearce and Tjalling Koopmans for a discussion of the problems involved in deciding how fast the country should move to the "optimum." One example of the use of the discount rate as a determinant of the rate of approach to the optimum golden age is contained in Hamada (1969).

Recalling that by definition $k'-D'=1$, we obtain the simpler form:

$$(14)\qquad (f')\left(1-\epsilon_{f'k}\cdot\epsilon_{ke}\cdot\frac{d}{e}\right)=g$$

where $\epsilon_{f'k}=f''k/f'=1/\sigma(1-f'k/f)<0$.

In an optimum state, under our assumptions about the shape of the equilibrium locus, we obtain the well-known result that the country should be a lender or borrower as the interest rate on capital at the zero debtor position exceeds or is less than the growth of population, i.e., as $f'|_{k=e}\gtrless g$. However, equation (13) implies that in general an open country should *not* save so as to equate the domestic interest rate with the growth rate of the effective labor force, and that the saving rate should generally *not* be equated with the output elasticity of capital or the share of capital under competition. The optimum values are related to the usual golden rule values in the following way:

$$f'[k^*]\lesseqgtr g,\qquad s^*\gtrless f'^*e^*/z^* \text{ as } (\epsilon_{k^*e^*})d^*\lesseqgtr 0$$

where the asterisk represents an optimum value. Thus a country which finds its optimum long-run equilibrium, as determined by equation (13), at a position as a net borrower ($d^*>0$) with foreign capital in less than perfectly elastic supply ($\epsilon_{k^*e^*}>0$), should drive its interest rate to less than the usual golden rule level and save more than usual golden rule implies.

In the special case in which $f'|_{k=e}=g$, the country's international balance of indebtedness is a matter of indifference, for equity and debt are perfect substitutes and the costs, in terms of foregone consumption, of increasing equity through domestic saving, g, are just equal to the costs of borrowing abroad, f'. This possibility exists for a whole range of equity labor ratios. In these circumstances the country could also afford to exercise any of its preferences for national, as opposed to foreign, ownership at no cost to itself in foregone consumption.

The intuitive explanation for these results revolves around the basic difference between equity accumulation in an open economy and capital accumulation in a closed economy, namely the difference between national product and domestic product. Each increase in per capita equity leads to some *net* increase in the capital-labor ratio and a corresponding increase in the per capita domestic product, equal to the marginal product of capital multiplied by the *net* change in the capital-labor ratio. However, increases in per capita national product are equal to increases in per capita equity multiplied by the marginal product of capital and will exceed increases in domestic product when equity and debt are substitutes. In the worst case—perfect substitutability—local accumulation simply buys up foreign capital,[27] without causing a decline in the interest rate, leading to no change in domestic product but a rise in national product. In the less extreme case of imperfect substitutability, there is a secondary beneficial effect on national as opposed to domestic product: the increase in the net capital-labor ratio decreases the interest rate on the foreign debt and, therefore, payments on debt. This effect on national product is positive in the case of a borrower, negative in the case of a lender.

In summary, benefits from accumulation as measured by the increment to national product from a small increase in the equity-labor ratio tend to be relatively constant when equity and debt are good substitutes. To the extent the two are not perfect substitutes, equity accumulation faces diminishing returns. However, even in this case diminishing returns to national equity are mitigated if the country is a borrower, since foreign debt holders

[27] The result is independent of the sign of D', so debt and equity need not be substitutes.

suffer a decrease in interest rates. In both cases there is an increase in the ratio of per capita national product to domestic product. Finally, in the special case in which equity accumulation actually stimulates foreign lending, diminishing returns to the nation are mitigated still further. Each increment in per capita equity induces greater foreign lending, a still larger increase in the capital-labor ratio, and a still greater increase in domestic product. With wages rising and the share of capital going to domestic capitalists also rising (recall that we assume $\epsilon_{ke}<1$), national product rises even faster and diminishing returns to the nation's equity are reduced still further.

The cost of obtaining these benefits or increases in per capita national product is, of course, the additional saving necessary to obtain the increase in per capita equity. Considering only balanced growth paths, the additional saving per capita which is necessary to maintain a small increase in the equity-labor ratio is simply equal to g, multiplied by the increase in equity per capita.[28] As the increase in per capita equity approaches zero, we obtain the marginal cost of steady-state accumulation which is simply g. Per capita consumption will increase if the increase in national product exceeds the increase in saving; maximum consumption is obtained by equating costs and benefits as shown mathematically in equations (13) and (14). There, benefits are set on the left-hand side, costs on the right. Since, as described above, benefits to a borrowing country exceed the interest rate, accumulation should be carried on to a point at which $g>f'$. The less elastic the supply of foreign capital and the more elastic the marginal product (the smaller σ), the greater should be the difference.

Graphically the rationale behind the solution is shown in discrete form by Figure 2, which omits the lower tier of Figure 1 and which is not drawn exactly to scale. An increase in the saving rate increases the long-run equilibrium level of per capita equity from Oe_1 to Oe_2. The corresponding increase in the capital-labor ratio is from Ok_1 to Ok_2 and the increase in domestic product is equal to $f'dk/de$ (not shown). However, as described above, the gain to the home country from an increase in saving is not the increase in domestic product but the increase in national product. Thus from $f'dk/de$ we must subtract the net increase in debt service payments. The change in debt service is shown in Figure 2 by the difference between the two rectangles $f'[k_1]k_1e_1A$ and $f'[k_2]k_2e_2D$. In turn this difference is equal to the sum of $f'[k_1]k_1e_1A-Ck_2e_2B+Cf'[k_2]DB$. The algebraic sum of the first two terms is the discrete representation of the reduction in interest payments to foreigners due to home country investors buying up debt ($f'[k]d'$ in equation (13)). The last term is the discrete representation of the secondary effect mentioned earlier, namely the decline in debt service payments due to the fall in the interest rate ($-f''k'd$ in equation (13)).

As argued above, in long-run equilibrium the additional saving necessary to maintain the new long-run capital-labor and equity-labor ratios is just equal to g, multiplied by the change in the equity-labor ratio. Thus the additional cost of foregone consumption which is required to increase per capita national product by the above net amount is the portion of increased national product which must be saved to maintain the new level of per capita equity, namely $g_2e_2e_1g_1$.

Figure 2 also graphs the function $f'\cdot(1-\epsilon_{f'k}\epsilon_{ke}(d/e))$ (from (14)) against k (and implicitly e) and shows that the optimum level of the capital-labor ratio is k_3. It is also easy to see that the country in

[28] See the paper by Neher and the author.

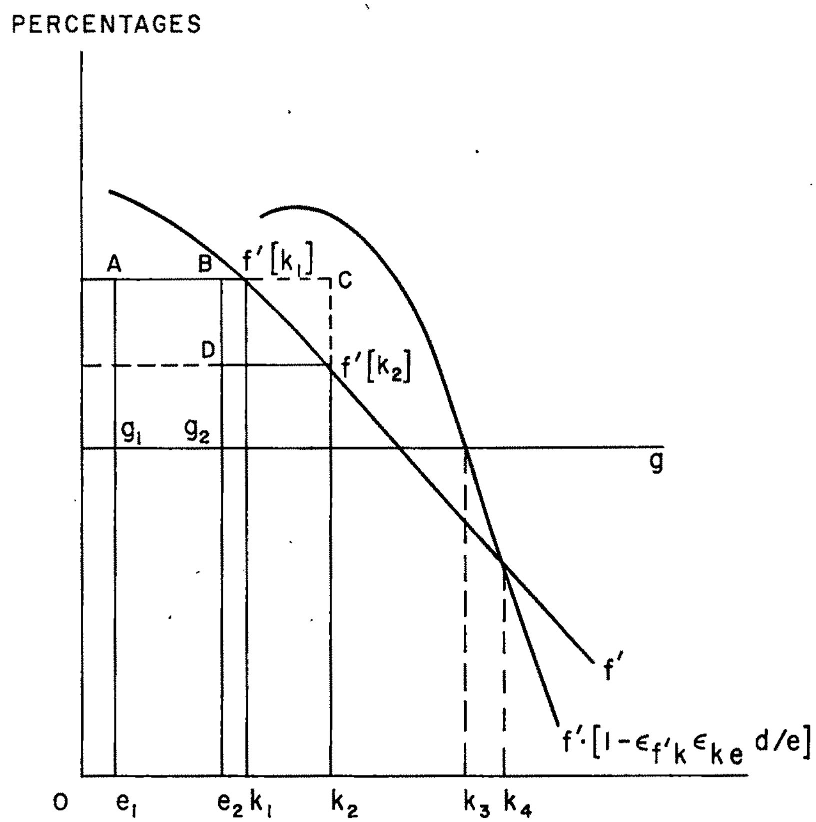


FIGURE 2

question is a borrower at k_3 by evaluating the left side of equation (14) at k_3. Since $f'\cdot(1-\epsilon_{f'k}\epsilon_{ke}(d/e))$ exceeds $f'[k_3]$ and $\epsilon_{f'k}\epsilon_{ke} \leq 0$, d must be positive, implying the country is a net borrower.

IV. Optimal Taxation Policies and Accumulation

Following Kemp's work, it is well known that a country may gain by imposing a tax on the return from international lending. This result may be easily seen by dropping the earlier assumption that foreigners and nationals receive the same rate of interest. In this case

$$z = y - id \tag{15}$$

where i= the rate of interest on international lending and supply of funds $d=k-e=D[k-e/k, i]$. Since the effective interest rate on debt can be held constant through manipulation of the taxes on service payments, the choice of optimal policies can be thought of as choosing the level of per capita equity which, given a fixed level of total interest payments on debt, maximizes national consumption, while also choosing the optimum amount of foreign investment and the corresponding debt service payments through variation in taxation on debt services. A global optimum is reached when

1) equity is expanded to the point where costs, i.e., the additional saving per man, just equal the benefits, i.e., addition to per capita national production, and

2) where the marginal returns on capital and debt are equal. Otherwise per capita national product could be ex-

panded by increasing the capital stock through increased debt or vice versa. This is equivalent to the optimum tariff in international trade theory.

The first equality is obtained mathematically by differentiating steady-state consumption $((1-s)z)$ with respect to equity, holding i fixed, and noting that $\partial k/\partial e = 1+\partial D/\partial e$. We obtain

$$f' = g - (f' - i)\partial D/\partial e \tag{16}$$

The sign of the last term depends on the partial effect of equity accumulation on lending, holding the interest rate constant. When foreign capitalists are indifferent to their relative share of the domestic capital stock, the last term vanishes and the condition reduces to the well-known golden rule. On the other hand, suppose that foreign capitalists feel that domestic accumulation decreases the risk of expropriation. Assume also that i, the average cost of foreign capital, differs from f', the marginal return to capital, a divergence which is clarified below. In that case expansion of domestic equity should proceed beyond the golden rule level. As shown by equation (16), a marginal increase in steady-state accumulation beyond the golden rule will induce still more foreign loans at less than the domestic cost of capital in terms of foregone consumption, namely, g.

The second equality is obtained mathematically by differentiating equation (15) partially with respect to debt, holding e constant and noting $\partial k/\partial d = 1$. We obtain

$$\frac{\partial z}{\partial d} = f' - i - \frac{\partial i}{\partial d} d = 0 \tag{17}$$

$$f' = i(1 + \epsilon_{id}),$$

where $\epsilon_{id} = (\partial i/\partial d)d/i$

The rule suggests borrowing up to the point at which the return from an additional unit of per capita debt, f', equals the marginal cost of debt, which in turn is equal to the interest rate on the last unit of lending plus the addition to interest costs on all previous debt.

Notice that under the assumptions of this paper, even a small country borrowing in the world market is likely to face a rising supply curve of foreign capital over much of its borrowing stage ($\epsilon_{id} > 0$). Even though additional borrowing may not affect world interest rates, any increase in the country's borrowing or foreign lending will influence the debt-equity ratio, and therefore the risk of expropriation. Thus even a small country must pay for greater borrowing with higher interest costs unless equity represents nearly all of the capital stock. In keeping with this reasoning, the term ϵ_{id} should not be interpreted as a pure partial elasticity, but an elasticity taking into account the need to compensate investors for any rise in risks associated with a rising debt-equity ratio.

In a borrowing country, equation (17) implies that $i = f' \cdot (1-t)$, where t is the tax rate equal to $\epsilon_{id}/(1+\epsilon_{id})$. Borrowing is restricted by levying a tax on returns to foreign capital at that rate. For a lending country, loans are restricted by levying a tax on the returns to foreign capital at the rate ϵ_{id}. In the case in which the optimum level of per capita equity occurs when the country is a small borrower or lender, which cannot affect the world interest rate, and in which foreigners are indifferent to the debt-equity ratio, there would, of course, be no tax. As discussed earlier, the optimum combination of debt and equity would then be a matter of indifference and the country could indulge in its preference for domestic capital at no cost in terms of foregone consumption. These results are, of course, variants of Kemp's basic proposition which incorporates domestic saving; they represent a more intuitive way of obtaining Hamada's result.

V. Conclusions

This paper has argued that the risk of national as well as individual default must be considered in international lending. It is then argued that one good measure of this risk is the ratio of national debt to the total capital stock, both from the point of view of making service payments out of income and from the point of view of visibility. This assumption—that the debt-equity ratio enters into lenders' calculations—provides a rationale for an upward sloping supply curve of foreign capital to small as well as large countries.

Under this assumption, national accumulation is doubly beneficial, for it raises national income both directly and indirectly through its effects on foreign investment and debt service payments. It follows that national accumulation should typically proceed beyond the levels suggested by the golden rule.[29]

These arguments apply to the case in which foreigners and nationals are taxed at similar rates on their investments. In the more general case the rate of return on debt may be held constant by variations in the taxation of profits on international investments. The argument of this paper does not modify the standard optimality conclusions for a lending country, namely that the marginal product of capital should be equated with the growth rate and that international loans should be taxed so as to equate the marginal returns on foreign and domestic investments. However, the results are very different than the standard conclusions in the case of a small borrowing country. In that case, domestic accumulation should typically proceed beyond the level suggested by the golden rule because of the effects on risk described above. Moreover, because of this risk even the small country might face an upward sloping supply curve of capital and would then find it advantageous, because of monopsony power, to tax foreign investment.

[29] Of course, this is under the assumption that increases in the debt-equity ratio stimulate foreign investment. If foreigners regard an increase as indicating greater risk, then accumulation should be stopped short of golden rule levels.

REFERENCES

G. Alter, "The Servicing of Foreign Capital Inflows by Underdeveloped Countries," in M. Ellis and H. Wallich, eds., *Economic Development in Latin America*, New York 1961.

D. Avramovic and R. Gulhati, *Debt Servicing Capacity and Postwar Growth in International Indebtedness*, Baltimore 1958.

——— and Associates, *Economic Growth and External Debt*, Baltimore 1964.

P. K. Bardhan, "Optimum Foreign Borrowing," in K. Shell, ed., *Essays on the Theory of Optimal Economic Growth*, Cambridge, Mass. 1967.

M. Bernstein, *Foreign Investment in Latin America: Cases and Attitudes*, New York 1966, 175–282.

J. Faaland, "Comment on Gulhati," in J. Adler, ed., *Capital Movements and Economic Development*, New York 1967.

R. Gulhati, "The Need for Foreign Resources, Absorbative Capacity and Debt Servicing Capacity," in J. Adler, ed., *Capital Movements and Economic Development*, New York 1967.

K. Hamada, "Economic Growth and Long-Term Capital Movements," *Yale Econ. Essays*, Spring 1966, *6*, 49–96.

———, "Optimal Capital Accumulation by an Economy Facing an International Capital Market," *J. Polit. Econ.*, July/Aug. 1969, Part II, *77*, 684–97.

J. Hanson and P. Neher, "The Neoclassical Theorem Once Again: Closed and Open Economies," *Amer. Econ. Rev.*, Sept. 1967, *57*, 869–78.

K. Inada, "On a Two-Sector Model of Economic Growth: Comments and a Generalization," *Rev. Econ. Stud.*, June 1963, *30*, 119–27.

H. Johnson, "Some Aspects of the Theory of

Economic Policy in a World of Capital Mobility," in T. Bagiotti, ed., *Essays in Honor of Mario Fanno*, Vol. II, Padova 1966, 345–59.

M. Kemp, *The Pure Theory of International Trade*, New York 1964.

———, *The Pure Theory of International Trade and Investment*, Englewood Cliffs 1969.

C. Kindleberger, *American Business Abroad*, New Haven 1969.

T. Koopmans, "On the Concept of Optimal Economic Growth," Cowles Foundation disc. pap. no. 163, Yale Univ. 1963.

G. Meier, *International Economics of Development*, New York 1968, ch. 5.

R. Nurkse, *Equilibrium and Growth in the World Economy*, Cambridge, Mass. 1961, 134–50, 259–78.

I. F. Pearce, "The End of the Golden Age in Solovia," *Amer. Econ. Rev.*, Dec. 1962, *52*, 1088–97.

A. Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, New York 1937.

H. Uzawa, "On a Two-Sector Model of Economic Growth," *Rev. Econ. Stud.*, Oct. 1961, *29*, 40–47.

R. Vernon, *How Latin America Views the U.S. Investor*, New York 1965.

On Biased Technological Progress

By Akira Takayama*

There has been an extensive study in the literature on the estimate of technological progress using an aggregative model. A typical assumption in this context is that of Hicks-neutral technological progress.[1] However, in theoretical models, the favorite assumption is that of Harrod-neutral technological progress, which probably reflects the desire of researchers to characterize the steady state.[2]

Obviously, however, there should be no a priori reason why technological progress should be either Hicks neutral or Harrod neutral. There have recently been interesting studies by Robert Resek, Paul David and Th. van de Klundert, and Ryuzo Sato (1970), and others to determine the bias of technological progress.[3]

The purpose of this note is to propose simple formulas which can be and are subsequently used to determine and estimate the type of bias of technological progress. My estimates indicate that technological progress in the United States for the period of 1909–60 could be Solow labor-saving, and that it has been neither Hicks neutral nor Harrod neutral. The non-Hicks neutrality confirms the results of Resek, David and van de Klundert, Sato (1970), and others. However, unlike most studies on the topic, I do not assume that the production function is a priori of the factor-augmenting type. Moreover, I do not, unlike David and van de Klundert, assume that the elasticity of factor substitution is a priori constant. In the last section, I investigate whether the elasticity of factor substitution has changed from period to period.

I. Model

In this section we obtain formulas which will be useful in determining the direction of technological progress. We begin with the usual production function which utilizes labor (L) and capital (K) under constant returns to scale[4]

$$Y = F(L, K, t) = Lf(k, t) \tag{1}$$

* Professor of economics, Purdue University. The idea of this paper was first presented at the University of Birmingham, England, in May 1965. I am indebted to Osamu Takenaka of Purdue University for computational help, and for useful comments and suggestions. I am also indebted to Tom Takuma of the University of Hawaii for computational help in an earlier version of the paper. K. Rao Kadiyala, Sheng Cheng Hu, George Borts, and the referee also provided me with useful comments.

[1] See, for example, Robert Solow, Benton Massel, Irving Kravis, Kenneth Arrow et al., John Kendrick and Ryuzo Sato, Phoebus Dhrymes, Leif Johansen, Mordecai Kurz and Alan Manne, Michio Morishima and Mitsuo Saito, and Charles Ferguson.

[2] See fn. 10.

[3] Resek pointed out the weakness involved in the test of Hicks neutrality in the well-known study by Solow. He then proposed his own test of Hicks neutrality, and, using it, refuted the Hicks neutrality for the *U.S.* economy. For a short summary of his test, see my fn. 9. Sato (1970) also refuted the Hicks neutrality of *U.S.* technological progress by using his own test. He then went on to estimate various parameters of the production function under Hicks nonneutrality. He assumed that technological progress is of the factor-augmenting type.

[4] A popular practice in the recent literature is to use the factor-augmenting type of production function, $Y=F[A(t)L, B(t)K]$, instead of (1). This may reflect the researchers' desire to incorporate technological progress embodied in the factors into their analysis (see, for example, David and van de Klundert). It goes without saying, however, that the above factor-augmenting type of production function is a special case of (1). As Hugh Rose, Sato and Martin Beckmann, and Sato (1970), have shown, (1) becomes a production function of the factor-augmenting type, if the elasticity of factor substitution is unaffected as long as income shares are fairly constant over time. In economies, such as that of the United States, in which factor shares are fairly constant over time, the use of the factor-augmenting type of production function then almost amounts to assuming the constancy of the elasticity of factor substitution σ, which itself is a testable hypothesis.

(where $k \equiv K/L$) in which disembodied technological progress is assumed. Define labor's relative share θ by[5]

$$\theta \equiv F_L L/Y \tag{2}$$

Clearly, capital's relative share is equal to $(1-\theta)$. Then, as is well known, we can obtain the following equations from (1) and (2)

$$\hat{Y} = \theta\hat{L} + (1-\theta)\hat{K} + \phi \tag{3}$$

$$\theta\hat{F}_L + (1-\theta)\hat{F}_K = \phi \tag{4}$$

where $\phi \equiv F_t/Y$ (the "rate of technological progress") and the hat denotes the proportional rate of change over time (for exexample, $\hat{Y} \equiv \dot{Y}/Y$). Denote the marginal product of labor and capital by $\mu(k, t)$ and $\nu(k, t)$, respectively.[6] Define the marginal rate of substitution of the two factors by

$$\omega(k, t) \equiv \mu(k, t)/\nu(k, t) \tag{5}$$

The elasticity of factor substitution (σ) is then defined by[7]

$$\sigma \equiv \frac{\omega}{\omega_k k} \tag{6}$$

Thus we obtain, from $\omega = \omega(k, t)$,

$$\hat{k} = \sigma\hat{\omega} + \lambda \tag{7}$$

where

$$\lambda \equiv -\sigma\omega_t/\omega \tag{8}$$

We say that technological progress is *Hicks neutral*, *Hicks capital-saving*, and *Hicks labor-saving* according to whether $\omega_t = 0$, $\omega_t > 0$, and $\omega_t < 0$, respectively.[8] Therefore, technological progress is (Hicks) neutral, capital-saving, and labor-saving according to whether $\lambda = 0$, $\lambda < 0$, and $\lambda > 0$. I shall refer to λ as the *Hicks index*.[9]

Since $\omega \equiv F_L/F_K$, we have $\hat{\omega} = \hat{F}_L - \hat{F}_K$, which combined with (4) yields

$$\hat{\omega} = (\phi - \hat{F}_K)/\theta = (\hat{F}_L - \phi)/(1-\theta) \tag{9}$$

Define H by

$$H \equiv \lambda - (1-\sigma)\phi/\theta \tag{10}$$

Then using (3), (7), (9), and (10), we obtain

$$\hat{Y} - \hat{K} = \sigma\hat{F}_K - \theta H \tag{11}$$

We say that technological progress is *Harrod neutral*, *Harrod capital-saving*, and *Harrod labor-saving* according to whether the output-capital ratio (Y/K) is kept constant, increases, or decreases, when the "profit rate" F_K remains constant. As is well known, and as we can easily see from (11), technological progress is (Harrod) neutral, capital-saving, and labor-saving according to whether $H = 0$, $H < 0$, and $H > 0$. H is termed the *Harrod index*.[10]

[5] F_L denotes $\partial F/\partial L$. Similarly $F_K \equiv \partial F/\partial K$ and $F_t \equiv \partial F/\partial t$.

[6] Needless to say, we have $\mu(k, t) = f(k, t) - kf_k(k, t)$ and $\nu(k, t) = f_k(k, t)$ where $f_k \equiv \partial f/\partial k$.

[7] As is well known, $\sigma > 0$ since $\omega_k > 0$, where $\omega_k \equiv \partial\omega/\partial k$. (Similarly ω_t denotes $\partial\omega/\partial t$.) In general, $\sigma = \sigma(k, t)$ and it is not constant. However, the class of production function in which σ is constant (called *CES*) plays an important role in economics, especially in empirical studies. A great interest was aroused by Arrow et al. For a useful survey of the *CES* production functions, see Mark Nerlove.

[8] Hence (as can easily be seen from (7)) technological progress is Hicks neutral, Hicks capital-saving, and Hicks labor-saving according to whether the capital-labor ratio (k) remains constant, decreases, or increases when ω is kept constant. ω is equal to the wage-rental ratio under the competitive situation.

[9] Assume perfect competition so that the wage-rental ratio q is equal to ω. Suppose that $\lambda = 0$ and $\sigma =$ constant. Then from (7), we have $k = aq^\sigma$ where a is some positive constant. Plot k vis-à-vis q in the scatter diagram. If we do not get such a relation which is stable, then we reject the hypothesis that technological progress is Hicks neutral, provided that $\sigma =$ constant. Hence if we cannot find such a stable relation, then we may conclude that this is due to a change in λ; i.e., we reject the hypothesis that technological progress is Hicks neutral. This is the method used by Resek. Sato (1970, pp. 191–92) pointed out some weaknesses of Resek's test and proposed his own test. His test, however, also indicates Hicks nonneutrality for the *U.S.* economy.

[10] Equation (11) shows why the assumption of Harrod neutrality is favored by many theorists. Let an *equilibrium growth path* be defined as the path in which the profit rate is constant ($\hat{F}_K = 0$) and capital grows at a constant rate ($\hat{K} =$ const.). Assume $\dot{K} = sY$ where s denotes the constant propensity to save. Then from (11), H must be equal to zero along the equilibrium growth path. If the growth rate of labor is given exogenously

Next we obtain the following relation from (7) and (9).

$$\hat{k} = \sigma(\phi - \hat{F}_K)/\theta + \lambda \tag{12}$$

If labor's relative share (θ) is constant so that $(1-\theta)$ is also constant, then from the definition of $(1-\theta)$ we have $\hat{Y}-\hat{K}-\hat{F}_K=0$. Combining this with (10), (11), and (12), we obtain

$$\lambda = (1-\sigma)\hat{k} \tag{13}$$

Therefore, in an economy (such as the *U.S.* economy) in which θ is fairly constant and $\hat{k}\neq 0$, the assumption of Hicks neutrality implies $\sigma=1$, the Cobb-Douglas production function. Therefore, it may not be surprising that Solow had a good fit with the Cobb-Douglas production function, since he accepted Hicks neutrality.[11]

Next, use (3) to obtain

$$\hat{Y} - \hat{L} = (1-\theta)\hat{k} + \phi \tag{14}$$

Then from (14), (7), and (9), we obtain

$$\hat{Y} - \hat{L} = \sigma\hat{F}_L + S \tag{15}$$

where

$$S \equiv (1-\theta)\lambda + (1-\sigma)\phi \tag{16}$$

Technological progress is said to be *Solow neutral*, *Solow capital-saving*, and *Solow labor-saving* according to whether the output-labor ratio (Y/L) remains constant, decreases, or increases when F_L is kept constant. As can be seen easily from (15), technological progress is (Solow) neutral, capital-saving, and labor-saving according to whether $S=0$, $S<0$, and $S>0$. S is termed the *Solow index*.[12]

It may be interesting to observe

$$S = (\lambda - \theta H) \tag{17}$$

which can be obtained from the definition of H and S. Equation (17) clarifies the relationships among the three indices. For example, it shows that if $\lambda>0$ and $H<0$, then $S>0$, and that if any two indices are zero, then the third index is zero (which, in view of (16), occurs only when $\sigma=1$, or when $\phi=0$).

II. Estimates[13]

The formulas which are most relevant for the present estimates are equations (7), (11), and (15). Assuming perfect competition so that $w=F_L$ and $r=F_K$, where w and r denote the real wage rate and real rent, respectively, (7), (11), and (15) are rewritten as

$$\hat{k} = \sigma\hat{q} + \lambda, \quad \text{where } q \equiv w/r \tag{18}$$

$$\hat{x} = \sigma\hat{r} - \theta H, \quad \text{where } x \equiv Y/K \tag{19}$$

$$\hat{y} = \sigma\hat{w} + S, \quad \text{where } y \equiv Y/L \tag{20}$$

The method of our study should now be clear. For example, if we attempt by regression to fit a straight line between $\hat{y}$ and $\hat{w}$, it is possible that the fit may not be good at all, for, in general, σ and S may change over time. But if we do get a good fit, then we may have reasonable grounds

and is constant, then in the "steady-state path" ("golden age path") in which $\hat{K}=\hat{L}$, $\hat{F}_K=0$ if and only if $H=0$. Note also that if $H=0$ and $\sigma\neq 1$, then $\lambda\neq 0$ in view of (10). Hence in order to obtain the steady-state path with $H=0$, and yet to insist on Hicks neutrality in view of the empirical claim by Solow, some economists are often led to assume that $\sigma=1$.

[11] This more or less confirms Sato's result (1970, pp. 285–90) which he obtained by using a more complete but much more tedious method than the present one. For the constancy of θ, see Lawrence Klein and Richard Kosobud, Sidney Weintraub, and Arthur Grant.

[12] As is well known, it has been shown that Solow neutrality for all t is equivalent to assuming the capital augmenting type of production function (see Sato and Beckmann). Also Harrod neutrality for all t is equivalent to assuming the labor augmenting type of production function (see Hirofumi Uzawa). Hicks neutrality for all t is equivalent to writing the production function in the form of $Y=A(t)G(L,K)$, (see Uzawa and Watanabe). Some economists have used the Cobb-Douglas form (see, for example, Dhrymes, Kurz and Mann, Morishima and Saito); then the three neutralities become indistinguishable.

[13] Like many other studies, our estimates are for the nonagricultural private sector of the *U.S.* economy. We shall use John Kendrick's data adopted in Sato's study (1970). Like Sato, we take the period of 1909 to 1960. See his table on pp. 202–03.

to suspect that σ is constant and that technological progress shows a definite Solow bias or neutrality. Needless to say, the slope of such a line measures σ and the intercept of the line with the $\hat{y}$ axis gives the Solow index. On the other hand, if we do not get a good fit, then we can assert that technological progress does not show a definite Solow bias or neutrality, and/or that σ is not constant. In view of (18) and (19), we may also wish to fit $\hat{k}$ vis-à-vis $\hat{q}$ and $\hat{x}$ vis-à-vis $\hat{r}$.

As Sato (1968, p. 284) points out, there are reasonable grounds to believe that σ may fluctuate over time; for example, σ may rise in a recession. To avoid any such bias, we shall estimate on a five-year moving average. The least squares estimate for equation (20) is now recorded here.

$$(20') \quad \hat{y} = \underset{(0.0502)}{0.6381}\,\hat{w} + \underset{(0.00148)}{0.00755},$$

$$R = 0.8843, \quad F = 161.3774$$

This looks like a good fit. The elasticity of substitution σ is found significantly different from zero. It is interesting to note that our estimate of σ is remarkably close to the past estimates.[14] The Solow index, which is estimated to 0.0075, is highly significantly different from zero. Hence we may conclude that technological progress has been Solow labor-saving.[15]

Suppose that θ= constant, so that $\hat{Y}-\hat{L}-\hat{F}_L=0$. Then from (15) we immediately obtain $S=(1-\sigma)\hat{F}_L$, or $S=(1-\sigma)\hat{w}$. In the $U.S.$ economy, θ is fairly constant and on the whole w has been rising. Therefore, if we accept $\sigma<1$ (where σ does not have to be constant), then we obtain $S>0$. This also confirms our observation that technological progress has been Solow labor-saving in the United States.

Our estimates of the other two equations are not as good as the above. Equation (18) is estimated as

$$(18') \quad \hat{k} = \underset{(0.0679)}{0.2446}\,\hat{q} + \underset{(0.00215)}{0.00760},$$

$$R = 0.4730, \quad F = 12.9686$$

The fit is rather poor. From this, we may conclude that any definite pattern of Hicks bias or neutrality is not apparent for the $U.S.$ economy. The interested reader can see this more dramatically by drawing the scatter diagram of $\hat{k}$ vis-à-vis $\hat{q}$.

However, it is interesting to observe that both the intercept and the slope of the regression line were found to be significantly different from zero. This may suggest that λ (and possibly σ) fluctuate over the years (in view of the low R), yet stay constant on the "average."[16] In view of a positive

[14] Unfortunately, virtually all time-series studies which estimate the elasticity of substitution of the $U.S.$ aggregate production function assume the Hicks neutrality of technological progress. Yet their estimates closely resemble our estimate σ^*. For example, $\sigma^*=0.64$ in Kravis, $\sigma^*=0.58$ in Kendrick and Sato, and $\sigma^*=0.67$ in Ferguson. Arrow et al. obtained $\sigma^*=0.57$, where they assume Hicks neutrality in the end, but their estimation of σ does not depend on this assumption. David and van de Klundert (D-K) and Sato (1970) estimated σ by explicitly avoiding the neutrality assumption. D-K obtained the estimate of σ as 0.619 and 0.316 by using two different methods. Sato (1970) obtained $\sigma^*=0.53$. Both D-K and Sato (1970) assumed that technological progress is of the factor-augmenting type. Unfortunately the correlation coefficient of Sato's estimating equation is very low: i.e., $R=0.3689$ for his equation (36). See Sato (1970, p. 194).

[15] Needless to say, in a strict sense, it is not true that we have proved this statement. We can only state that a Solow labor-saving type of technological progress is consistent with the $U.S.$ data. However, we believe that we have examined such a consistency by use of a sharp test. In this connection, we may recall the well-cited "impossibility theorem" which proposes the impossibility of measuring the elasticity of factor substitution and the bias of technological progress simultaneously. This theorem appears to be correct. But in any empirical study it is hard to accept a hypothesis; hence it would be difficult to accept the hypothesis that technological progress has a certain bias. Usually, we can only hope to create a test which is sharp enough so that we are more inclined to accept the hypothesis tested.

[16] Notice also the difference in the estimate of σ between (20′) and (18′), which may suggest the fluctuation of σ over the years. However, it is possible (and perhaps

estimate of the intercept, we may also conclude that technological progress has been Hicks labor-saving ($\lambda > 0$). Incidentally, if we accept that $\sigma < 1$ and that θ has been fairly constant with $\hat{k} > 0$ for the *U.S.* economy, then we can also conclude from (13) that $\lambda > 0$.

Finally, we record the regression estimate of equation (19).

$$(19') \quad \hat{x} = \underset{(0.0936)}{0.6461}\,\hat{r} + \underset{(0.00237)}{0.00504},$$

$$R = 0.7170, \quad F = 47.6046$$

The fit is not as good as that of (20), so the result here has to be interpreted with more reservation. However the fit is not as bad as that of (18). The estimate of the coefficient of $\hat{r}$, which measures the elasticity of substitution σ, is approximately equal to 0.65, and it was found that σ is significantly different from zero. This estimate of σ is very close to the one obtained earlier from (20′). The estimate of the intercept is again close to zero, but positive. If we accept that it is positive, then technological progress has been Harrod capital-saving.[17]

Suppose that θ has been more or less constant. Furthermore, suppose that we accept the constancy of σ from (20′). Then, in view of the aforementioned result (see fn. 4) by Rose, and Sato and Beckmann, we may conclude that the factor augmenting type of technological progress is consistent with the *U.S.* data.[18] In the next section, we shall test the constancy of σ further.

Therefore, as a conclusion, we only state that there is some indication that the *U.S.* economy has been labor-saving in the sense of Solow, and also perhaps in the sense of Hicks. There is no definite indication of Harrod neutrality as theoreticians may hope to be the case. Rather, there are some indications of Harrod capital-saving technological progress.[19]

III. Further Considerations

In connection with equation (20′) we observed that σ is significantly different from zero, and that its estimate is close to the estimates obtained in some other studies. However, it is also well known that there is no concensus with regard to the estimated value of σ even on the aggregate level.[20] Moreover, it is again well known that the elasticity of factor substitution differs from industry to industry. Hence, if the composition of industries in the national economy changes in the course of economic growth, it may be natural to suppose that the value of σ changes over time. Moreover, as we noted in the previous section, there is some indication of the fluctuation of σ over the years. Also, Sato and Ronald Hoffman suggest that σ may

likely) that σ is indeed constant on the average, and that the fluctuation of λ has caused the bad fit of (18′) and a difference in the estimate of σ. In view of a rather successful fit of (20′) and the tests of the constancy of σ in the next section, we shall take this view. Thus $\sigma^* = 0.64$ in (20′) gives an estimate of σ for the entire period.

[17] Suppose that capital's relative share is constant so that $\hat{Y} - \hat{K} - \hat{F}_K = 0$. Then from (11), we immediately obtain $H = -(1-\sigma)\hat{F}_K/\theta$. Therefore if $\sigma < 1$ and $\theta =$ constant, $H \gtreqless 0$ according to whether $\hat{F}_K \gtreqless 0$ (or $\hat{r} \gtreqless 0$). There has been an increase in the value of r in the United States during the period of 1909–60, and this is consistent with the indication of Harrod capital-saving technological progress, if $\sigma < 1$. However, it is also known that there has been a considerable amount of fluctuation in r. Hence, it is also not surprising if we find no strong indication of Harrod capital-saving bias.

[18] If σ changes while θ is constant, then the technological progress cannot be of the factor-augmenting type. Suppose that the production function is written in the factor-augmenting form $Y = F[A(t)L, B(t)K]$; then it can be shown that $S = (1-\sigma)\hat{A}$, $-\theta H = (1-\sigma)\hat{B}$, and $\lambda = (1-\sigma)(\hat{A} - \hat{B})$, where $\hat{A} \equiv \dot{A}/A$ and $\hat{B} \equiv \dot{B}/B$. Hence, if σ and S are constant while λ is not, then $\hat{A}$ is constant but $\hat{B}$ is not. If σ and S are estimated, respectively, as 0.64 and 0.00755 (see equation (20′) and fn. 16), then $\hat{A}$ is estimated as 0.02, which resembles closely the 0.0193 in Sato (1970, p. 195).

[19] If technological progress has been Solow labor-saving and Harrod capital-saving, then in view of footnote 12, the aggregate production function can be of neither the (purely) capital-augmenting nor the (purely) labor-augmenting type. This does not preclude the possibility of the form $Y = F[A(t)L, B(t)K]$.

[20] See, for example, Nerlove.

increase over time due to the "enhancement of opportunities of factor substitution" (p. 458).

However, the study in the previous section suggests that although σ may fluctuate over the years, it is constant on the average. Hence, in this section, we divide the entire period into three "epochs," estimate the value of σ for each epoch, and then test whether σ has indeed changed over the epochs.[21] More specifically, following Murray Brown and Joel Popkin, I divide the entire period into three "technological epochs." That is, call the period 1911–18 "epoch I," the period 1919–37 "epoch II," and the period 1938–57 "epoch III." First, estimate the value of σ for each epoch. For this purpose, define the following dummy variables:

(a) D_{I}: $D_{\mathrm{I}}(t)=1$ for all t in epoch I and $D_{\mathrm{I}}(t)=0$ for all t in epochs II and III.
(b) D_{II} and D_{III} are defined analogously.
(c) $D_{\mathrm{I,II}}$: $D_{\mathrm{I,II}}(t)=1$ for all t in epochs I and II and $D_{\mathrm{I,II}}=0$ for all t in epoch III.
(d) $D_{\mathrm{II,III}}$ and $D_{\mathrm{I,III}}$ are defined analogously.

We are now ready to report the following estimates for equation (20).[22]

$$(21)\quad \hat{y} = \underset{(0.0995)}{(0.7636}D_{\mathrm{I}} + \underset{(0.0512)}{0.6196}D_{\mathrm{II,III}})\hat{w} + \underset{(0.00146)}{0.00767}, \quad R = 0.8879, \; F = 83.75$$

$$(22)\quad \hat{y} = \underset{(0.0583)}{(0.6551}D_{\mathrm{II}} + \underset{(0.0607)}{0.6185}D_{\mathrm{I,III}})\hat{w} + \underset{(0.00149)}{0.00764}, \quad R = 0.8852, \; F = 79.688$$

$$(23)\quad \hat{y} = \underset{(0.0644)}{(0.5756}D_{\mathrm{III}} + \underset{(0.0537)}{0.6697}D_{\mathrm{I,II}})\hat{w} + \underset{(0.00147)}{0.00787}, \quad R = 0.8904, \; F = 84.20$$

$$(24)\quad \hat{y} = \underset{(0.0993)}{(0.7585}D_{\mathrm{I}} + \underset{(0.0571)}{0.6489}D_{\mathrm{II}} + \underset{(0.0643)}{0.5748}D_{\mathrm{III}})\hat{w} + \underset{(0.00147)}{0.00789}, \quad R = 0.8934, \; F = 56.68$$

From these four equations, we can observe that the estimate of σ is around 0.6 for epochs II and III and somewhat higher than this for epoch I. The standard error of each estimate is always very low. The estimate of the intercept is about 0.007 to 0.008, which resembles closely the estimate in (20′), and shows only small variation.[23] Compared to (20′), the value of R does not show much improvement.

We now wish to test whether the value of σ has indeed shifted over the epochs. First we record the following estimates where (20) is now fitted epoch by epoch.[24]

[21] Sato and Hoffman tested the hypothesis that σ is constant for each year but changes over time according to $\sigma=a+bt$ ("shifting *CES* functions"). If we find that σ does not change over the epochs, then this provides some evidence against their hypothesis. Our conclusion in this section will thus be somewhat contrary to that of Sato and Hoffman.

[22] We also estimated similar regressions for equations (18) and (19). As in the case of (18′) and (19′), we again do not get good fits, and we shall thus not report our estimates.

[23] Whether or not the value of the intercept (i.e., the value of S) differs on the "average" from epoch to epoch is tested by using the dummy variable technique. It is concluded that there are no such significant changes. For example, we may record:

$$\hat{y} = \underset{(0.1106)}{(0.8280}D_{\mathrm{I}} + \underset{(0.0703)}{0.6133}D_{\mathrm{II}} + \underset{(0.1031)}{0.5293}D_{\mathrm{III}})\hat{w} + \underset{(0.00371)}{0.00492} + \underset{(0.00447)}{0.00442}D_{\mathrm{II}} + \underset{(0.00396)}{0.00447}D_{\mathrm{III}}$$

where $R=0.8897$, $F=34.43$. It was found that the intercepts with respect to D_{II} and D_{III} are *not* significantly different from 0, which can be shown to be equivalent to concluding the nonrejection of (H_0: $S_{\mathrm{I}}=S_{\mathrm{II}}$) and of ($H_0$: $S_{\mathrm{I}}=S_{\mathrm{III}}$), respectively. Here S_i denotes the value of S for epoch $i(i=\mathrm{I, II, III})$. Other studies on the variability of S_i using alternative methods of testing are left to the interested reader.

[24] A weakness of these estimates is obviously the decrease in the sample size compared to earlier estimates. Hence we may use (21)–(24) mainly to estimate the value of the σ_i, and (25)–(27) mainly to test the relative stability of the σ_i over the epochs.

$$(25)\qquad \hat{y}_{\mathrm{I}} = \underset{(0.1205)}{0.8280}\,\hat{w}_{\mathrm{I}} + \underset{(0.00278)}{0.00492},$$

$$R = 0.9420, \quad F = 47.2283$$

$$(26)\qquad \hat{y}_{\mathrm{II}} = \underset{(0.0619)}{0.6133}\,\hat{w}_{\mathrm{II}} + \underset{(0.00196)}{0.00934},$$

$$R = 0.9234, \quad F = 98.3170$$

$$(27)\qquad \hat{y}_{\mathrm{III}} = \underset{(0.1106)}{0.5293}\,\hat{w}_{\mathrm{III}} + \underset{(0.00325)}{0.00939},$$

$$R = 0.7484, \quad F = 22.9143$$

Let σ_i be the value of σ for epoch $i(i=\mathrm{I, II, III})$.[25] To examine the relative stability of σ, we test the null hypothesis

$$H_0: \sigma_i = \sigma_j (i \neq j;\ i, j = \mathrm{I, II, III})$$

against

$$H_1: \sigma_i \neq \sigma_j (i \neq j;\ i, j = \mathrm{I, II, III})$$

In view of (20), this can be tested by computing

$$(28)\qquad t_{ij} = (\overset{*}{\sigma_i} - \overset{*}{\sigma_j})/s_{\sigma_i^* - \sigma_j^*}, \qquad i \neq j; \quad i, j = \mathrm{I, II, III}$$

which are distributed as "Student's" t. Here σ_i^* and $s^2_{\sigma_i^* - \sigma_j^*}$, respectively, stand for the regression estimates of σ_i and the estimated variance of $(\sigma_i^* - \sigma_j^*)$.[26] The values of the t_{ij} can easily be computed as

$$(29)\qquad t_{\mathrm{I,II}} = 1.747, \quad t_{\mathrm{II,III}} = 0.683, \quad t_{\mathrm{I,III}} = 1.836$$

Referring to the t-table with the relevant degrees of freedom, we may then conclude:

(i) Either of (H_0: $\sigma_{\mathrm{I}} = \sigma_{\mathrm{II}}$) and ($H_0$: $\sigma_{\mathrm{I}} = \sigma_{\mathrm{III}}$) can be rejected at the 5 percent significance level but *not* at the 2.5 percent significance level.

(ii) (H_0: $\sigma_{\mathrm{II}} = \sigma_{\mathrm{III}}$) *cannot* be rejected either at the 5 percent or the 2.5 percent (or even at the 10 percent) significance level.

So much for the "pair-wise" stability of σ over the epochs. Next we may test the "overall" stability of σ: i.e., let the null hypothesis be

$$H_0: \sigma_{\mathrm{I}} = \sigma_{\mathrm{II}} = \sigma_{\mathrm{III}}$$

The test can be carried out by computing the F-ratio, from which we can conclude that H_0 *cannot* be rejected at the 5 percent significance level.[27]

In view of the above pair-wise test and the overall test, we may conclude that there is no overwhelming evidence of changes in the value of σ from epoch to epoch,[28] although there may be a slight indication of a change between the first and the second epochs.[29] In other words, our estimates show some indication that the *U.S.* economy, especially for epochs II and III, has changed its industrial composition so as to maintain a constant value of σ. The reason why this has happened might be a topic for future research.

[25] The estimates of σ_{I}, σ_{II}, and σ_{III} from (25)–(27) are 0.8280, 0.6133, and 0.5293, and it was found that each σ_i, $i=\mathrm{I, II, III}$, is highly significantly different from 0.

[26] The computational procedure is described in Bernard Ostle, p. 205.

[27] See Ostle, pp. 201–05, for example. The value of F in equation (8.122) of Ostle is computed for the present case as $F=2.03$. Then, referring to the F-table, observe that $F_{0.95}^{(2,41)} \cong 3.23 > 2.03$, where 2 and 41 stand for the degrees of freedom for the present study. Concerning the variability of σ between epoch I and the other two epochs, there is an apparent contradiction between the present conclusion and the above conclusion based on the t-test. However, this is due to whether the data for all three epochs are pooled or not. In each of the pair-wise t-tests, the data for two epochs are relevant, while in the overall F-test, the data for all three epochs are necessary.

[28] Or, more precisely, our grouping of periods by epoch à la Brown and Popkin does not create highly significant evidence for the changes in σ.

[29] The relative stability of σ can also be examined by using the dummy variable technique. For example, if we estimate $\hat{y} = (a_{\mathrm{I}}D_{\mathrm{I}} + a_{\mathrm{II}}D_{\mathrm{II}} + a_{\mathrm{III}}D_{\mathrm{III}}) + \{b + (b_{\mathrm{II}}D_{\mathrm{II}} + b_{\mathrm{III}}D_{\mathrm{III}})\}\hat{w}$ (or $\hat{y} = a + \{b + (b_{\mathrm{II}}D_{\mathrm{II}} + b_{\mathrm{III}}D_{\mathrm{III}})\}\hat{w}$ under the assumption of $a_{\mathrm{I}} = a_{\mathrm{II}} = a_{\mathrm{III}}$), and if b_{II} and b_{III} are *not* significantly different from 0, we cannot reject either (H_0: $\sigma_{\mathrm{I}} = \sigma_{\mathrm{II}}$) or ($H_0$: $\sigma_{\mathrm{I}} = \sigma_{\mathrm{III}}$). On the other hand, if, for example, b_{II} is significantly different from 0, then we can reject the hypothesis (H_0: $\sigma_{\mathrm{I}} = \sigma_{\mathrm{II}}$).

REFERENCES

K. J. Arrow, H. B. Chenery, B. Minhas, and R. M. Solow, "Capital Labor Substitution and Economic Efficiency," *Rev. Econ. Statist.*, Aug. 1961, *43*, 225–50.

M. Brown and J. Popkin, "A Measure of Technological Change and Returns to Scale," *Rev. Econ. Statist.*, Nov. 1962, *44*, 402–11.

P. A. David and T. van de Klundert, "Nonneutral Efficiency Growth and Substitution Between Capital and Labour in the U.S. Economy, 1899–1960," *Amer. Econ. Rev.*, June 1965, *55*, 357–94.

P. A. Diamond and D. McFadden, "Identification of the Elasticity of Substitution and the Bias of Technical Change: An Impossibility Theorem," working pap. no. 62, Inst. Bus. Econ. Res., Univ. California, Berkeley, Mar. 1965.

P. J. Dhrymes, "A Comparison of Productivity Behavior in Manufacturing and Service Industries," *Rev. Econ. Statist.*, Feb. 1963, *45*, 64–69.

C. E. Ferguson, "Substitution, Technical Progress, and Returns to Scale," *Amer. Econ. Rev. Proc.*, May 1965, *55*, 296–305.

A. Grant, "Issues in Distribution Theory: The Measurement of Labor's Relative Share, 1899–1929," *Rev. Econ. Statist.*, Aug. 1963, *45*, 273–97.

J. R. Hicks, *The Theory of Wages*, 2d ed., London 1963.

W. P. Hogan, "Technical Progress and Production Functions," *Rev. Econ. Statist.*, Nov. 1958, *40*, 407–11.

L. Johansen, "A Method of Separating Effects of Capital Accumulation and Shifts in Production Function upon Growth in Labor Productivity," *Econ. J.*, Dec. 1961, *71*, 775–82.

J. W. Kendrick, *Productivity Trends in the United States*, Princeton 1961.

——— and R. Sato, "Factor Prices, Productivity and Economic Growth," *Amer. Econ. Rev.*, Dec. 1963, *53*, 974–1003.

L. R. Klein and R. F. Kosobud, "Some Econometrics of Growth: Great Ratios of Economics," *Quart. J. Econ.*, May 1961, *75*, 173–98.

I. B. Kravis, "Relative Income Shares in Fact and Theory," *Amer. Econ. Rev.*, Dec. 1959, *49*, 917–49.

M. Kurz and A. S. Manne, "Engineering Estimates of Capital-Labor Substitution in Metal Machining," *Amer. Econ. Rev.*, Sept. 1963, *53*, 662–81.

B. F. Massel, "Capital Formation and Technological Change in the United States Manufacturing," *Rev. Econ. Statist.*, May 1960, *42*, 182–88.

M. Morishima and M. Saito, "A Dynamic Analysis of the American Economy, 1902–1952," *Int. Econ. Rev.*, May 1964, *5*, 125–64.

——— and ———, "An Economic Test of Sir John Hicks's Theory of Biased Inventions," in J. N. Wolfe, ed., *Value, Capital, and Growth, Papers in Honour of Sir John Hicks*, Edinburgh 1968.

M. Nerlove, "Recent Empirical Studies of the CES and Related Production Functions," in M. Brown, ed., *The Theory and Empirical Analysis of Production*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth, 31*, New York 1967, 55–122.

B. Ostle, *Statistics in Research*, 2d ed., Ames 1963.

R. W. Resek, "Neutrality of Technical Progress," *Rev. Econ. Statist.*, Feb. 1963, *45*, 55–63.

H. Rose, "The Condition for Factor-Augmenting Technical Change," *Econ. J.*, Dec. 1968, *78*, 966–71.

R. Sato, "The Estimation of Biased Technical Progress and Production Function," *Int. Econ. Rev.*, June 1970, *11*, 179–208.

———, *The Theory of Economic Growth*, (in Japanese), Tokyo 1968.

——— and M. Beckmann, "Neutral Inventions and Production Functions," *Rev. Econ. Stud.*, Jan. 1968, *35*, 57–66.

——— and ———, "Neutral Inventions and Production Functions: An Addendum," *Rev. Econ. Stud.*, July 1968, *35*, 366.

——— and R. F. Hoffman, "Production Functions with Variable Elasticity of Factor Substitution: Some Analysis and Testing," *Rev. Econ. Statist.*, Nov. 1968, *50*, 453–60.

R. M. Solow, "Technical Change and the Aggregate Production Function," *Rev. Econ. Statist.*, Aug. 1957, *39*, 312–20.

H. Uzawa, "Neutral Inventions and the Stability of Growth Equilibrium," *Rev. Econ. Stud.*, Feb. 1961, *28*, 117–24.

——— and T. Watanabe, "A Note on the Classification of Technical Innovations," *Econ. Stud. Quart.*, 1961, *22*, 68–71.

S. Weintraub, *A General Theory of the Price Level, Output, Income Distribution, and Economic Growth*, Philadelphia 1959.

The Sizes and Types of Cities

By J. V. HENDERSON*

This paper presents a general equilibrium model of an economy where production and consumption occur in cities. The paper solves for equilibrium and optimum city sizes, discussing under what situations the equilibrium size differs from the optimum. Optimum city sizes are defined as those which maximize potential welfare of participants in the economy. Equilibrium city sizes are determined by the location or investment decisions of laborers and capital owners, each attempting to maximize their own perceived welfare.

Some of the basic concepts underlying the model are contained in the following propositions. We observe population agglomeration or cities because there are technological economies of scale in production or consumption and because these activities are not space or land intensive (relative to agriculture). Scale economies may occur at the final output level, at the marketing level, or at the intermediate input level, such as in transportation systems or capital and labor market development.

Given the existence of scale economies, what limits city size? The following argument is developed by Edwin Mills, and I utilize his basic argument in this paper. Mills assumes urban production of traded goods to occur in a central business district (*CBD*). In addition to traded goods, housing is produced in the city and workers commute to the *CBD* from their sites surrounding the *CBD*. As city size and the area devoted to housing increase spatially, the average distance a worker commutes necessarily increases as does congestion. That is, average per person commuting costs rise with city size. Efficient city size occurs where these increasing per person resource costs offset the resource savings due to scale economies in traded good production.

Why do cities vary in size? This question pertains basically to Section IV of paper, since in the main body of the paper cities will all be the same size and type. City sizes vary because cities of different types specialize in the production of different traded goods, exported by cities to other cities or economies. If these goods involve different degrees of scale economies, cities will be of different sizes because they can support different levels of commuting and congestion costs. But why do cities specialize?

Provided there are no positive production benefits or externalities from locating two industries together, locating the production of the two goods in the same city only works to raise total production costs. Laborers employed in the two industries contribute to rising per person commuting costs, but scale economy exploitation occurs only with labor employment *within* each industry. If we locate the industries together, there are higher average per person commuting resource costs for a *given* level of scale economy exploitation or industry employment within either industry than if we locate the industries in separate cities. This is one reason why cities will tend to specialize in the production of different traded goods. To be weighed against the specialization advan-

* Assistant professor, Brown University. This paper was written while I was at Queen's University, Canada. I am indebted to Charles Upton and George Tolley for extensive comments on earlier versions of this paper and for many fruitful discussions on the topics and ideas presented in the paper. Patricia Munch, George Borts, Harry Johnson, and an anonymous referee also provided helpful comments on the paper.

tage are the transportation costs of trade between specialized cities. Goods such as retailing services are not traded between cities because of high transportation costs. Note that cities will probably specialize in bundles of goods, where, within each bundle, the goods are closely linked in production. They may use a common specialized labor force or a common intermediate input.

Throughout this paper, it is assumed that capital and labor are scarce resources available in fixed supply to the national economy. The economy is defined as a region or country within which these factors are perfectly mobile. Factors move between cities in the economy to equalize appropriate measures of factor rewards. The economy is situated on a flat featureless plain large enough so that land per se has zero opportunity cost and is never a scarce resource.

As stated above, I deal mainly with an economy where there is only one type of city. Each city produces and exports the same traded good at a fixed price to other regions or countries. In return the cities import another consumption good at a fixed price. In Section IV, I outline complications of the model presented in my thesis that incorporate multiple types of cities trading with each other in the same economy (where the terms of trade may be endogenous). Although these complications are interesting, they are not needed to develop the basic ideas in this paper.

I. The Model of a City

The model of a single representative city is presented in this section and solved for factor reward equations (which may refer to either or both factor prices and utility levels). Factor rewards will be a function of city employment of capital and labor and the fixed price of the city's export good. In Section III, the factor reward equations will be used to solve equilibrium and optimum city size for all cities in the economy subject to the economy endowment of resources. City size indicates the allocation of the economy's factors to each city, the number of cities, and the prevailing level of factor rewards which are equalized between cities (for equilibrium in factor markets). Given that there is only one type of city for most of the paper, city sizes will turn out to be all identical. We now turn to our representative city and develop the factor reward equations.

A. *Production Conditions*

Our representative city produces a traded good X_1 under conditions of increasing returns to scale, external to the firm but internal to the industry and city. These scale economies are responsible for the urban agglomeration discussed above. The industry production function is

$$X_1^{1-\rho_1} = L_1^{\alpha_1} K_1^{\beta_1} N_1^{\delta_1}, \tag{1}$$

$$\alpha_1 + \beta_1 + \delta_1 = 1, \quad 0 < \rho_1 < 1$$

where L_1, N_1, and K_1 are inputs of home or land sites, labor, and capital, respectively. The variable ρ_1 represents the degree of increasing returns to scale; hence $(\alpha_1+\beta_1+\delta_1)/(1-\rho_1)>1$. As stated above, X_1 is sold by the city at a fixed price set in national or international markets.

Under this externality specification (as explained in an article by John Chipman), the *firm* views itself as having a constant returns to scale production function. Therefore the private marginal product of, say, labor in the industry is $\delta_1 X_1/N_1$ rather than the social marginal product $[\delta_1/(1-\rho_1)]X_1/N_1$. This preserves exhaustion of firm revenue by factor payments. Atomistic competition is ensured since any entering firm benefits from the existing level of externalities or industry scale economies, i.e., firm size is unimportant in the model. Later in the paper I briefly discuss the fact that since social and private marginal

products differ, factor allocations may not be strictly Pareto optimal.

The second good produced is X_3, housing services for workers living in the city. Since housing services are a nontraded good, their price will vary with city size and will be determined in the model. The production function for housing is

$$(2) \qquad X_3 = N_3^{\delta_3} K_3^{\beta_3} L_3^{\alpha_3}, \qquad \delta_3 + \beta_3 + \alpha_3 = 1$$

The third good produced in the city is sites, an intermediate input in X_3 and X_1 production. In a spatial model, a site used in the production of housing is produced with an input of raw land and labor (time) inputs of commuting needed for travel to the *CBD* from a spatial location in the city. These commuting costs of producing sites escalate as city size increases and average commuting inputs or distance and congestion increase. In addition, increased use of sites in X_1 production competes with sites for residential use and therefore contributes to rising commuting distances and costs.

For mathematical simplicity without crucial omissions in economic reasoning, the spatial world is collapsed into a nonspatial world in this paper. Rather than explicitly having spatial dimensions or commuting in the paper, I simply assume sites are produced with labor inputs subject to decreasing returns to scale or rising per site labor inputs as city size increases. I hypothesize the model works qualitatively "as if" it were a spatial model. With no spatial dimensions, people will have identical housing consumption but the average resource costs of sites and housing will rise as city size increases. Also, there is no separate class of land owners in the model.[1]

Sites are produced with labor or commuting time inputs and used in X_1 and X_3 production. That is,

$$(3) \qquad (L_1 + L_3)^{1-z} \equiv L^{1-z} = N_0 \qquad z < 0$$

where L and N_0 are sites and labor inputs (raw land is not specified separately in (3) since its opportunity cost is zero). The variable z represents the degree of decreasing returns to scale. (Furthermore z is assumed to increase in absolute value with city size. Specifically it is assumed $1/(1-z) = N^m$, $-1 < m < 0$ where N is city population. The reason for this assumption is essentially algebraic and is mentioned later.) The diseconomies of scale are assumed to be external to the individual, and so again while factor payments exhaust revenue, factor allocations to sites are not strictly Pareto optimal. This problem is commented on briefly later. Intuitively the externality exists because when a laborer enters a city he imposes higher average commuting costs on other city inhabitants (see James Buchanan and Charles Goetz).

Given the production functions for the three goods produced in the city, the production side of the model is completed by the resource and intermediate input employment equations where N and K are city population and capital stock.

[1] The other crucial aspect of the commuting phenomenon in a spatial model is land rents. Residential location theory as developed by Richard Muth and Mills tells us there is a spectrum of commuting costs and land rents in a city. Land rents act as a rationing device so that people who live nearer the *CBD* and experience lower commuting costs pay higher rents to offset their cost advantage relative to those further from the *CBD*. The actual land itself involves no resource costs if its opportunity cost is zero. The land rents are a transfer from renter to landowner reflecting the relative "scarcity" of a location. In a nonspatial world there is no role for a rationing device or spectrum of land rents and landowners. Rising resource costs of commuting are captured but the location scarcity principle is not represented. However, given land rents are essentially a transfer from renter to landowner, our results concerning equilibrium and Pareto optimum city size are unaffected. But to the extent that rising land rents induce further substitution away from homesite inputs in housing and X_1 production, the resource costs of the commuting phenomenon are "under-represented" in our model.

$$N_0 + N_1 + N_3 = N \qquad (4)$$
$$K_1 + K_3 = K$$
$$L_1 + L_3 = L$$

B. *Consumption Conditions*

To close the model, consumption conditions must be specified in order to derive the demand equations for three consumer goods. In addition to goods produced in their own city, X_1 and X_3, city inhabitants consume a good X_2 imported from other economies at fixed price q_2. Consumers have identical tastes and maximize logarithmic linear utility functions subject to their income and prices q_1, q_2, and q_3. Income spent in the city and city demand for X_1, X_2, and X_3 is determined as follows.

Laborers live in the cities where they work and spend their income. Capital owners are not constrained to live in the city where their capital rentals are earned. They may live in the countryside, in other cities, or in other countries. Cities may be net borrowers or lenders with respect to the proportion of capital rentals earned versus spent in the city. Given these problems and varieties of situations, two alternative polar assumptions are made. These assumptions play a crucial analytical role later in the paper.

Assumption A. All capital owners live in the cities of this economy and also work as laborers. For simplicity it is assumed capital ownership is evenly divided among laborers. If the cities in our economy have the same K/N ratio in production (which they will as long as they are identical or until Section IV), they are neither *net* borrowers nor net lenders. However, they may be gross borrowers or lenders since capital owners need not invest in the city they live in.

Assumption B. Capital owners are a separate group of people who do not work as laborers. They avoid the high cost of living or housing in cities (see below) by living in the countryside or other countries. (Since we have fixed the supply of capital to the economy in this paper, living in the countryside makes more sense.) *No capital rentals are spent in the cities of this economy.*

Summarizing the consumption conditions, we have individuals maximizing utility, $U = x_1^a x_2^b x_3^c$ where x_i is individual consumption of x_i, subject to, for Assumptions A and B, respectively, either $y = p_N + p_K\ K/N$ or $y = p_N$ where p is factor price. From this optimization process, we may obtain expressions for the indirect utility function of an individual and the aggregate demand equations for the city (the sum of individual demands). Where Y is city income which equals yN, these expressions are

$$X_1^C = aY/q_1, \quad X_2^C = bY/q_2, \quad X_3^C = cY/q_3 \qquad (5)$$

$$U = a^a b^b c^c y q_1^{-a} q_2^{-b} q_3^{-c} \qquad (6)$$

Equation (6) for the indirect utility function is used extensively throughout the paper. In (5), the superscript $C(P)$ refers to goods consumed (produced) in the city. This distinction is crucial for the balance of trade equation for the city $X_1^P q_1 - X_1^C q_1 = X_2^C q_2 + k p_K K$ where $k = 1$ under Assumption B and $k = 0$ under Assumption A if cities are not net lenders or borrowers.

II. Solution of the Model for a City

From the consumption and production equations of the model, city output, exports, factor prices, and the price of sites and housing can be solved for in terms of city employment of capital K and labor N and the fixed prices of traded goods, q_1 and q_2. City employment of capital and labor will be determined in Section III when we solve for city size. In this section we simply explore how equilibrium factor prices, housing prices, and, in particular, utility levels vary as we vary K and N. To solve for equilibrium movements of factor and housing prices as we vary city employment

of K and N, we solve our model of a city. We combine the supply and demand side by combining the full employment equations (4) for our representative city with the private marginal product equations, determining factor prices in the X_3, X_1, and L industries, the consumer demand equations for X_1 and X_3 from equation (5), and the cost functions derived from (1), (2), and (3). The method of solution is detailed in my thesis as is the derivation of all equations in the model. Given the solutions in terms of K and N for housing prices, wage rates (p_N), and capital rentals (p_K), we may substitute these variables into (6) to solve for utility levels as a function of city employment of capital and labor.

Assumption B. In determining equilibrium city size, we must consider the location decision of a laborer. A laborer will choose to live in the city that he perceives as maximizing his utility. From equation (6) his utility is a function of just two variables: q_3, the price of housing or the city cost of living and p_N, the city wage rate. Both of these as explained can be solved for in terms of city employment of K and N. Equation (6) also defines the welfare of laborers and is used in solving for optimum city sizes. Below we present the expression for utility of laborers in terms of K and N called U_N which we will use to help solve for both equilibrium and optimum city sizes under Assumption B.

For capital owners, the distinction between variables governing investment decisions and those reflecting the benefits of such decisions will be crucial. Capital owners do not have to live in cities where their capital is employed and under Assumption B do not live in cities at all. In determining equilibrium city size, we assume capital owners invest to maximize capital rentals. Under Assumption B (only), capital rentals also reflect the welfare of investment decisions of capital owners. Since their cost of living or housing is independent of their investment decisions and the size and cost of living in cities in this economy, p_K is the only variable in equation (6) in terms of the benefits to capital owners of their investment decisions.

Therefore, under Assumption B, we use an expression for the utility of laborers, U_N, to analyze the location decisions of laborers in solving for equilibrium city size. We use p_K to analyze the investment decisions of capital owners. We also use these same variables to solve for optimum city sizes.

Assumption A. Now since laborers are also capital owners and all factor payments are spent in the cities of our economy, things are not so simple. There are two basic decisions determining equilibrium city size. One is an investment decision and the other a location decision. Since laborers can invest their capital in any city in the economy, not just the one they live in, their investment decision is divorced from their location decision of which city to live in. Therefore when we examine the capital market in determining equilibrium city size, we will assume laborers seek to maximize capital rentals which then must be equalized between cities for capital markets to be in equilibrium.

In making location decisions, laborers seek to maximize utility as a function of the variables in (6): namely q_3, the price of housing, and $p_N+\overline{p_K}\ \overline{K/N}$, the per laborer income, where p_K is exogenous to the location decision[2] and $\overline{K/N}$ is the fixed amount of capital *owned* by each laborer. Per laborer utility is then

[2] If one assumes capital is physically tied to the owner as for, say, a small business owner, then he will move his business to maximize an index of utility—the capital rentals deflated by a cost of living. For example a small business owner in New York will demand a higher return on his capital than if he lived in Albany, simply due to cost of living differences.

$$U = a^a b^b c^c (p_N + \overline{p_K}\,\overline{K/N}) q_1^{-a} q_2^{-b} q_3^{-c}$$

For expositional simplicity we split per laborer utility levels into the sum of two parts:

$$U_N = a^a b^b c^c p_N q_1^{-a} q_2^{-b} q_3^{-c}$$

plus

$$U_K = a^a b^b c^c \overline{p_K}\,\overline{K/N} q_1^{-a} q_2^{-b} q_3^{-c}$$

The level U_N is utility from labor income and U_K is utility from capital income. Again, since as explained above, p_N, p_K, and q_3 can be expressed in terms of city employment of K and N, U_N and U_K can similarly be expressed in terms of K and N.

In summary, to solve for equilibrium city size under Assumption A we depict investment decisions as investors or laborers seeking to maximize capital rentals or p_K. Location or migration decisions are depicted by laborers seeking to maximize the sum of U_N and U_K where $\overline{p_K}$ in U_K is exogenous to the location decision.

In determining optimum city sizes, we are not concerned with separate investment and location decisions. Instead we are concerned with the simultaneous determination of p_K and utility levels that maximizes utility or welfare of laborers in the economy. Since laborers are the only participants in our economy, we simply seek to maximum their welfare which is the sum of U_N and U_K where p_K is no longer exogenous to the location or any other decision. The precise meaning of these statements will become apparent below; but, to repeat, we seek to maximize U_N plus U_K, *given* the determination of U_N, U_K, and p_K through simultaneous location and investment of labor and capital in cities in the economy.

The following equations are given and used in subsequent analysis of city size. As explained above, housing prices and factor prices in terms of city employment of K and N are obtained by combining the city full employment equations, the industry marginal product equations, the consumer demand equations, and industry cost functions. The exponents of the equations contain production and consumption parameters including z, the degree of decreasing returns to scale in homesite production (where $1/(1-z)=N^m, 0>m>-1$). For example, the expression for p_K depicts the equilibrium movement of capital rentals or private marginal product equations (for example, $p_K=q_1B_1\,X_1/K_1$) in the X_1 and X_3 industries, where the equilibrium movement of p_K is determined by the production and consumption conditions of our model. With nonconstant returns to scale characterizing production functions, equilibrium p_K is a function of the city K/N ratio, the scale of output, and the degrees of increasing (ρ_1) and decreasing (z, where $1/1-z=N^m$) returns to scale.

Under Assumption A, in addition to p_K, we present expressions for the utility U_N from wage income (p_N) and the utility U_K from capital income ($p_K\ \overline{K/N}$).[3] As explained above these are obtained by substituting expressions for p_N, P_K, and q_3 into (6). We write the equations in na-

[3] The expressions for p_N and q_3 under Assumption A are

$$\begin{aligned} \log p_N &= \log (C_N q_1) + N^m \frac{\alpha_1}{\rho_1 - 1} \log t \\ &\quad + \frac{\alpha_1(1-N^m)-\rho_1}{\rho_1 - 1} \log N \\ &\quad + \frac{-\beta_1}{\rho_1 - 1} \log K/N \\ \log q_3 &= \log (C_q q_1) + N^m \frac{(\alpha_1 - \alpha_3(1-\rho_1))}{\rho_1 - 1} \log t \\ &\quad + \frac{\beta_3 - \beta_1 - \beta_3\rho_1}{\rho_1 - 1} \log N \\ &\quad + \frac{(\alpha_1 - \alpha_3(1-\rho_1))(1-N^m)-\rho_1}{\rho_1 - 1} \log K/N \end{aligned}$$

where C_N and C_q are constant.

$$(7)\quad \log U_N = \log(W_N q_2^{-b} q_1^{1-c-a}) + N^m \frac{(\alpha_1(1-c) + c\alpha_3(1-\rho_1))}{\rho_1 - 1} \log t + \frac{(-\beta_1 - c\beta_3 + c\beta_1 + c\beta_3\rho_1)}{\rho_1 - 1} \log(K/N) + \left(\frac{(\alpha_1(1-c) + c\alpha_3(1-\rho_1))(1-N^m) - \rho_1(1-c)}{\rho_1 - 1}\right) \log N$$

$$(8)\quad \log p_K = \log(C_K q_1) + N^m \frac{\alpha_1}{\rho_1 - 1} \log t + \frac{1-\rho_1-\beta_1}{\rho_1 - 1} \log K/N + \frac{\alpha_1(1-N^m) - \rho_1}{\rho_1 - 1} \log N$$

$$(9)\quad \log U_K = \log \overline{K/N} + \log(W_K q_2^{-b} q_1^{1-c-a}) + N^m\left(\frac{\alpha_1(1-\rho_1) + c\alpha_3(1-\rho_1)}{\rho_1 - 1}\right) \log t + \left(\frac{1-\rho_1-\beta_1+c\beta_1+c\beta_3(\rho_1-1)}{\rho_1 - 1}\right) \log K/N + \left(\frac{(\alpha_1(1-c) + \alpha_3 c(1-\rho_1))(1-N^m) - \rho_1(1-c)}{\rho_1 - 1}\right) \log N$$

tural logarithmic form as (7), (8), and (9). The coefficients W_N, C_K, and W_K are constants defined in my thesis and are not relevant to our discussion. In addition,

$$t = \left(\alpha_1 + \alpha_3 \frac{c}{1-c}\right)^{-1} \cdot \left(\alpha_1 + \delta_1 + (\alpha_3 + \delta_3)\frac{c}{1-c}\right) > 1$$

Note that in (9), we distinguish between the fixed $\overline{K/N}$ in ownership describing the quantity of capital owned by each individual and the variable city K/N ratio in production which is determined in the model. (In this paper since there is only one type of city, in the following it will turn out all cities are identical in size and economic characteristics and, hence, $\overline{K/N}$ will equal K/N in production.) By inspection it can be seen $\partial U_N/\partial(K/N) > 0$ and $\partial p_K/\partial(K/N)$, $\partial U_K/\partial(K/N) > 0$ if $\rho_1 + \beta_1 < 1$. That is, normal factor ratio effects on factor rewards prevail unless the degree of increasing returns to scale ρ_1 is very large.

Under Assumption B, we are only concerned with the expressions for utility from spending labor income and for capital rentals. The expressions for U_N and p_K are identical to (7) and (8) except the constants W_N, C_K, and t are replaced by W'_N, C'_K, and s^{-1} where $s = [(1-\delta_3 c - \alpha_3 c)\alpha_1 \cdot (\alpha_1 + \delta_1)^{-1} - \alpha_3 c] < 1$. It can be shown that normally $s^{-1} > t$.

A. *Utility and Capital Rental Paths*

The equations presented above are a function of the K/N ratio in the city, a measure of scale of city output or N, a variety of production consumption parameters, and prices, q_1 and q_2. We want to see how factor rewards vary with city size or N so that in Section III we may determine equilibrium and optimum city sizes or N. To do this, we isolate the scale effect from the factor ratio effect and any effect of changing q_1 and q_2. We take the derivative of the above equations with respect to N, or city size, holding K/N, q_1 and q_2 constant. We will show later our analysis is neutral or unaffected by changes in K/N, q_1, and q_2. Using the derivatives of factor rewards, the values that factor rewards assume with different city sizes will be summarized in *factor reward paths*. The derivatives of (7), (8), and (9) are shown in (10), (11), and (12). These equations are

$$(10)\qquad \frac{\partial p_K}{\partial N} = N^{m-1} p_K \frac{\alpha_1 m}{\rho_1 - 1}\left[\log t - 1/m + \frac{\alpha_1 - \rho_1}{\alpha_1 - m} N^{-m} - \log N \right]$$

$$(11)\qquad \frac{\partial U_K}{\partial N} = U_K N^{m-1}\left(\frac{m(\alpha_1 - c\alpha_1 - c\alpha_3(\rho_1 - 1))}{\rho_1 - 1}\right)$$
$$\cdot\left[\log t - \frac{1}{m} + \frac{(1-c)(\alpha_1 - \rho_1) + c\alpha_3(1-\rho_1)}{m(\alpha_1 - c\alpha_1 - c\alpha_3(\rho_1 - 1))} N^{-m} - \log N \right]$$

$$(12)\qquad \frac{\partial U_N}{\partial N} = \frac{U_N}{U_K}\frac{\partial U_K}{\partial N}$$

analyzed in my thesis and here we just summarize the relevant results.

The sign of the derivatives is given by the sign of the expressions in the square brackets in each equation. If N is small, the expressions are all positive indicating that initially capital rentals and utility levels rise as city size rises. As N increases, either the derivatives remain positive or become negative, depending on whether the signs of the third terms in the square brackets are positive or negative.

A sufficient condition (necessary for p_K) that *both* capital rentals and utility levels rise to maximum and then decline[4] or that the derivatives eventually become negative is that $\alpha_1 \geq \rho_1$. The variable α_1 represents the intensity with which the resource input, land sites, is used in X_1 production, and ρ_1 is the degree of scale economies in X_1 production. If $\alpha_1 \geq \rho_1$, factor rewards attain a maximum and decline because the benefits of agglomeration (ρ_1) are eventually offset by scale diseconomies in site production where the level of site production rises as α_1 rises. This net change in efficiency will be reflected in factor prices p_K and p_N which will reach a maximum and then decline. Moreover, consumption benefits such as U_N or U_K of spending marginal products are *further* limited because to obtain U_N or U_K, factor prices are deflated by q_3^{-c}. The cost of housing q_3 rises with city size as sites become more expensive. This effect is reflected in the c and α_3 parameters in equation (11) where they represent the share of housing in consumption and of homesites in housing production.

If $\alpha_1 < \rho_1$, either the utility levels may reach a maximum while capital rentals rise indefinitely or both utility and capital rentals may rise indefinitely. Due to space limitations, we only discuss the case where capital rentals and utility levels both rise to a maximum and then decline.

The changes in factor rewards with city size are illustrated in Figure 1. We draw utility levels and capital rentals on the same diagram for convenience although they are measured in different units on the vertical axis. The city sizes where factor rewards are maximized, $N(p_K^*)$, $N(U_N^*)$, and $N(U_K^*)$, may be determined by solving equations (10)–(12) equated to zero. The maximum $N(U_N^*) = N(U_K^*) \equiv N(U_K^*, U_N^*)$ but it can be shown that $N(U_K^*, U_N^*) < N(p_K^*)$. This is not surprising since, say U_K^* is p_K^* *further* deflated by q_3^{-c}, the price of housing which rises with city size.

Two further comments are in order. Given $\alpha_1 \geq \rho_1$ from equations (7)–(9), the following normal results can be shown to

[4] The reason why we specified $z(1/(1-z)=N^m)$ as variable, occurred because if z is fixed the derivatives in (10)–(12) are either always positive or always negative. That is, factor rewards never climb to a maximum and then decline, the situation we are interested in. Alternative to z varying, is ρ_1 varying or both varying. Our choice is arbitrary.

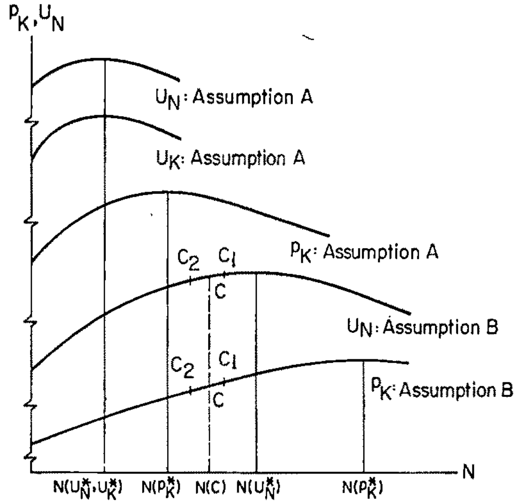


Figure 1. Utility and Capital Rental Paths Under Assumptions A and B

prevail: $\partial p_K/\partial(K/N)$, $\partial U_K/\partial(K/N)<0$, $\partial U_N/\partial(K/N)>0$, and all derivatives with respect to q_1 are positive. A change in K/N or q_1 shifts the factor reward paths up or down as indicated by the sign of the derivations. Further, from (10)–(12) equated to zero, it can be seen that the city sizes where the factor reward paths attain a maximum are invariant with respect to changes in K/N and q_1.

To examine factor reward paths under Assumption B, *log t* is replaced by $-log\ s$ in equations (10) and (12). The above discussion in terms of relative values of α_1 and ρ_1 and the shape of the utility and capital rental paths applies directly. However, given the values of t and s cited above, because $-log\ s$ has replaced the smaller *log t* in equations (10)–(12), the city sizes where U_N and p_K are maximized are larger than under Assumption A. This is not surprising. Under Assumption B, no capital rentals are spent in cities nor thus devoted to increasing the demand for housing produced and consumed in the city. Since the amount of housing and hence number of sites produced relative to X_1 is smaller, the cost of sites rises more slowly to offset the benefits of agglomeration. The variables U_N and p_K achieve a maximum at a larger city size under Assumption B.

III. City Size

In this section, the utility and capital rental paths derived above are used to solve for city size. Optimum city size is found by maximizing welfare of participants in the economy. Equilibrium city size is determined given atomistic optimization behavior in the investment and location decisions of capital owners and laborers.

To initiate the process of city formation, we start with one city in the economy producing X_1 and then increase the size of the economy. This does not mean we have a growth model per se since we have no savings behavior, technological change, etc. It is an artificial and simple method of solving for city size. However it does yield solutions for optimum city sizes and does serve to reveal the possibility of inadequate functioning of the market forces and signals that determine equilibrium city sizes. In this paper we do not discuss how the economy K/N ratio is determined. Presumably there are underlying growth relationships in the economy and well-defined saving and investment behavior. With only one type of city under discussion in this paper as explained above, changes in the K/N ratio do not affect the shapes of the factor reward paths and as we will see below do not affect when new cities form. We just assume a K/N ratio which may or may not change (given macro-economic conditions in the economy) and a growing economy.

A. *Stability Conditions*

Two or more cities cannot be of sizes such that they are both on the rising part of the utility and capital rental paths. Only if there is a *single* city, can a city be on the rising part of both factor reward

paths. For example, in Figure 1 under Assumption B, suppose there are two cities of size $N(C)$. A random movement of a small amount of capital and labor from one city to the other would move us from C to C_1 in the receiving city and from C to C_2 in the losing city. Since U_N and p_K both rise in the receiving city this will induce further movements of factors to the receiving city. That is, the initial equilibrium is unstable with respect to factor movements. However it is stable to have two cities on the rising part of the U_N path and falling part of the p_K path. *Throughout the paper we rule out unstable solutions as possible optimal* (and of course equilibrium) *solutions*. The above stability arguments are properly developed in my thesis.

B. *Optimum City Size*

We now determine optimum city size. The discussion serves only to solve for optimum city size and does not reflect or indicate behavior on the part of capital owners and laborers. There is initially one city in the economy and the economy is growing. We want to know when it is optimal to form a second, then a third, etc., city. For the initial discussion the K/N ratio and q_1 are held constant by assumption. In solving for optimum city size Assumptions A and B play a crucial role. We solve first for Assumption A.

Under Assumption A, all capital owners work as laborers in this economy and hence endure the cost of living in the city when spending their income. Although capital owners invest to maximize capital rentals, we are concerned with their utility from spending capital rentals (and labor wages) when solving for optimum city size. As explained above, to solve for optimum city size we maximize laborers' utility from wage and rental income or we maximize the vertical sum of the U_N and U_K paths.

In Figure 2, holding K/N constant as city size grows beyond $N(U_N^*, U_K^*)$, the

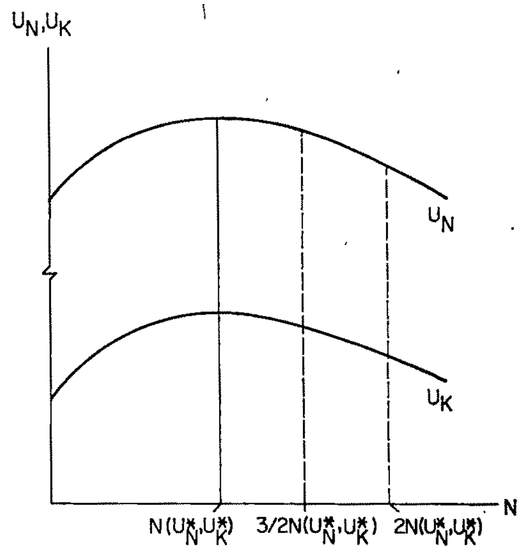


FIGURE 2. OPTIMUM CITY SIZE: ASSUMPTION A

city size of maximum U_N and U_K, a second city should form when N equals twice $N(U_N^*, U_K^*)$. The new solution then has two cities of size $N(U_N^*, U_K^*)$, resulting in stability in factor markets, equalization of factor rewards between cities, and full employment of factors in the economy. If two cities formed before $2N(U_N^*, U_K^*)$, resulting in city sizes less than $N(U_N^*, U_K^*)$ on the rising part of the factor payment paths, stability would not prevail in factor markets. Note that dividing utility into the sum of utility from capital and labor income presents no problems in analysis since these paths attain a maximum at the same point and hence their sum attains a maximum at $N(U_N^*, U_K^*)$. As the two cities of size $N(U_N^*, U_K^*)$ continue to grow, a third city of size $N(U_N^*, U_K^*)$ should form from the two cities when they reach size $3/2N(U_N^*, U_K^*)$. In general, a $n+1$ city should form when the n cities reach size $(n+1/n)\ N(U_N^*, U_K^*)$. If $n \rightarrow \infty$, which will be called the *large sample case*, city size will approach $N(U_N^*, U_K^*)$ where U_N and U_K are maximized. From equations (11) and (12), N equals $N(U_N^*, U_K^*)$ can be solved from:

$$(13)\quad \log t - 1/m + \frac{(1-c)(\alpha_1 - \rho_1) + c\alpha_3(1-\rho_1)}{m(\alpha_1 - c\alpha_1 - c\alpha_3(\rho_1 - 1))} N^{-m} - \log N = 0$$

A change in the K/N ratio as the economy grows would not affect city size. Regardless of K/N, U_N and U_K always attain a maximum at the same city size and hence optimal city sizes as well as equation (13) are unaffected.

In the discussion here and in particular for the discussion of market equilibrium to follow, we are discussing abstract solutions in which new cities form instantaneously. In the real world this would not happen, of course. For example, assume a growing economy with one city. When it is appropriate to form a second city in the economy, the second city would start off very small growing over time until the two cities were the same size. Given in reality the nonmalleability of capital, the actual population decline in the first city when a second city forms might be very small. This would be particularly true if growth in the economy was accompanied by technological change that increased efficient city sizes. Then the first city might not decline in size, but the second city would grow more rapidly over time until the two cities were the same size.

Note finally the city size $N(U_N^*, U_K^*)$ indicates the *maximum benefits of scale economies or welfare for our economy*. The benefits of scale economies are limited by the costs of agglomeration or the rising costs of homesites or commuting in a spatial world. In a certain sense, at $N(U_N^*, U_K^*)$ we approach a constant returns to scale case in production. Doubling the size of the economy would only double the number of cities and would bring no further scale economy benefits.

Under Assumption B, if capital owners live in the countryside or abroad in other countries, the price they pay for housing is independent of the urban cost of living. Therefore they maximize utility by maximizing their only variable in equation (6), p_K. To solve for optimum city sizes the p_K and U_N paths are utilized in the same fashion as the paths for Assumption A. The difference here is that the points where capital owners want to form cities indicated by $N(p_K^*)$ in Figure 3 are *different* than the points where laborers want to form cities indicated by $N(U_N^*)$. How are these differences reconciled? Before proceeding we note as mentioned above that although U_N and p_K are measured in different units, we draw them on the same diagram. Hence the vertical axis measures capital rentals in dollars or utility levels in utility units.

If the initial city has reached twice $N(U_N^*)$ in Figure 3, laborers would be better off if two cities of size $N(U_N^*)$ formed and capital owners worse off. The size of the initial city should increase beyond twice $N(U_N^*)$ *if from the increase in earnings of capital owners we can compensate laborers for their loss in utility from not forming a second city*. In other words, we employ a Pareto optimality criterion to define optimal city sizes. Our criterion includes capital owners, whether they live in the econ-

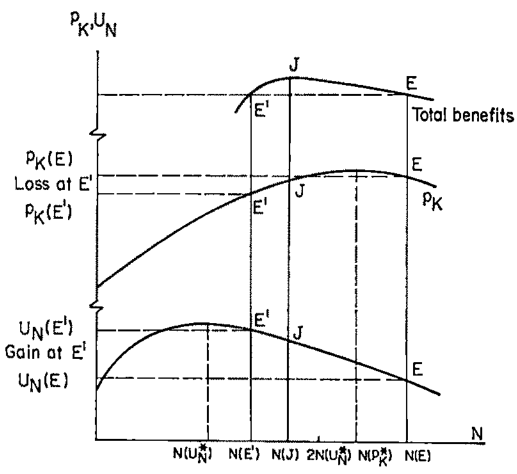


FIGURE 3. OPTIMAL CITY SIZE: ASSUMPTION B

omy (the countryside) or in other regions or other countries. Second, it is important to understand that our compensation mechanism is used only to depict a Pareto optimal solution as might be administered by an omniscient ruler. I do not envision groups of laborers and capital owners bribing each other to attain Pareto optimal solutions in the free market. Market behavior of capital owners and laborers and market solutions are discussed in the next section.

Suppose the initial city size moves beyond twice $N(U_N^*)$ to $N(E)$ where it is optimal to form a second city, yielding two cities of size $N(E')$ where $N(E) = 2N(E')$. At $N(E)$, we can no longer compensate laborers from the earnings of capital owners for not forming two cities of size $N(E')$. As illustrated in Figure 3, the loss to capital owners of two cities forming is $K(p_K(E) - p_K(E'))$ and the gain to laborers is $N(U_N(E') - U_N(E))$. The compensation that we could give to individual laborers from capital owners for not forming a second city is $M(K)$ and the compensation needed by a laborer for not forming a second city is $M(N)$ where

$$(14) \quad M(K) = K/N(p_K(E) - p_K(E'))$$

$$U_N(E') - U_N(E) = a^a b^b c^c q_1^{-a} q_2^{-b} q_3^{-c} M(N)$$

or

$$(15) \quad M(N) = a^{-a} b^{-b} c^{-c} q_1^a q_2^b q_3^c \cdot (U_N(E') - U_N(E))$$

From equation (6), the q_3^{-c} in (15) is used to deflate the compensation $M(N)$ where q_3 is the price of housing in city size $N(E)$. The variable $M(N)$ is the income subsidy to laborers needed to raise utility levels in city size $N(E)$ to those in the smaller city size $N(E')$. (Note that the calculation of p_N and q_3 and hence $U_N(E)$ will be affected by $M(N)$ since the demand for housing in city size $N(E)$ will rise if income is subsidized by $M(N)$.) At $N(E)$, two cities of size $N(E')$ form because $M(N) \geq M(K)$, where both $M(N)$ and $M(K)$ are measured in dollars.

After two cities of size $N(E')$ form, the economy continues to grow with additional optimal size cities forming via our compensation mechanism. Of particular interest is the large sample case where the number of cities is very large and hence when an additional city forms the changes in city size of existing cities are minimal. Since city size changes are minimal, the changes in U_N and p_K and the compensation that could be made from the earnings of capital owners and needed by laborers in equations (14) and (15) can be expressed in derivative form. In Figure 3, $N(J)$ is picked as the optimal point to form an additional city and optimal city size approaches $N(J)$. Note that $N(U_N^*) \leq N(J) \leq N(p_K^*)$. At J, $M(N) \geq M(K)$ or by directly substituting in (14) and (15)

$$(16) \quad M(N) = \left| a^{-a} b^{-b} c^{-c} q_1^a q_2^b q_3^c (\partial U_N/\partial N) \right| \geq \left| K/N(\partial p_K/\partial N) \right| = M(K)$$

From equation (15), $a^{-a}b^{-b}c^{-c}q_1^a q_2^b q_3^c$ converts $\partial U_N/\partial N$ to dollars. Substituting in $\partial U_N/\partial N$ and $\partial p_K/\partial N$ from equations (10) and (12) for Assumption B, (16) becomes equation (17). Solving (17) for N would yield $N = N(J)$, the optimum city size. A change in the K/N ratio would not affect optimum city size or equation (17), just as it would not affect $N(U_N^*)$ and $N(p_K^*)$.

Our compensation mechanism defines Pareto optimal solutions. Until $N(J)$ or optimal city size is reached, we can move city sizes towards $N(J)$ and take from the income of the gaining group of factors and compensate the losing group, such that both parties benefit by moving closer to optimum city size $N(J)$. Alternatively one can view the process as maximizing a hypothetical total benefit curve (or "the size of the pie" given that factor incomes are determined by marginal productivity con-

$$(17)\quad \left|\left[C'_N \frac{m(\alpha_1 - \alpha_1 c - c\alpha_3(\rho_1 - 1))}{\rho_1 - 1}\right] \cdot \left[-\log s - \frac{1}{m} + \frac{(1-c)(\alpha_1 - \rho_1) + c\alpha_3(\rho_1 - 1)}{m(\alpha_1 - c\alpha_1 - c\alpha_3(\rho_1 - 1))} N^{-m} - \log N\right]\right| \gtrless \left|C'_K \frac{\alpha_1 m}{\rho_1 - 1}\left[-\log s - \frac{1}{m} + \frac{\alpha_1 - \rho_1}{\alpha_1 m} N^{-m} - \log N\right]\right|$$

ditions). Total benefits are the "sum" (converted into common units) of utility levels plus capital rentals each weighted by factor endowments. Benefits are maximized where $N(a^{-a}b^{-b}c^{-c}q_1^a q_2^b q_3^c)\ \partial U_N/\partial N + K\,\partial p_K/\partial N = 0$ as in (16) at $N(J)$. Before $N(J)$, gains in capital income contributing to total benefits exceed losses in utility levels by moving towards $N(J)$. After $N(J)$, losses in utility levels contributing to total benefits exceed gains in capital rentals. We draw in the hypothetical total benefit curve in Figure 3. Optimal city size and maximum benefits of agglomeration occur at $N(J)$.

Three other points are of interest before we turn to market equilibrium solutions. First, since $N(U_N^*)$ under Assumption B lies beyond $N(U_N^*)$ for Assumption A as discussed in Section II, and since optimal city size lies beyond (rather than at) $N(U_N^*)$ under Assumption B, city sizes under Assumption B are larger than under Assumption A. Secondly, as should be obvious, under both Assumptions A and B if capital rental and utility paths rise indefinitely, a possibility mentioned in Section II, the optimum number of cities will be one.

Finally, before leaving the discussion on optimality, we mention a problem raised when we specified the production side of the model. The production of X_1 and L involved economies of scale external to the firm and hence, following Chipman, these industries should be respectively subsidized and taxed to correct for divergences of private from social costs. This problem is independent of our analysis, and from my thesis, it appears the effect of these taxes on city size is not large and one can assume in our discussion that they have been accounted for.

C. *City Formation and Size: A Market Economy*

We now solve for equilibrium city size in the economy. In our *initial naive* solution, the market economy is characterized by atomistic behavior of capital owners, firms, and laborers. For the initial discussion we deal only with Assumption B. The market behavior of factor owners is depicted by laborers moving between cities to maximize utility levels and capital owners investing to maximize capital rentals. Therefore we use the U_N and p_K paths to solve for equilibrium city size.

Initially it is the behavior of firms that determines city size and formation in our economy. Starting with a single city in the economy, a second city forms when a firm sees it is profitable to leave the first city, hire factors competitively, and set up a second city. However, because scale economies are external to the firm, individual firms act unaware of the scale economies that could accrue to their own size of operation. When they move to form a second city they hire an arbitrarily small amount of factors and initially set up an arbitrarily small firm size and city. (With external economies of scale and firms having linear homogeneous subjective production functions, firm size is indeterminate.)

In Figure 4, a firm can hire small

amounts of capital and labor away from the initial city when it reaches size $N(E)$. In the new arbitrarily small city of size N(small), the entrepreneur will *initially* operate with a lower K/N ratio, explaining the shifts in the U_N and p_K paths. (The firm could operate (inefficiently) without the paths shifting at N(small) when the initial city size is slightly larger than $N(E)$ at existing K/N ratios in production by paying laborers less than their marginal product and capital more than its marginal product.) Given the shifts in the paths and the difference in scale of operations, the entrepreneur can pay capital rentals and utility levels equal to or greater than competitive ones. The greater than specification allows him to earn profits for setting up the new city. These profits will encourage other firms to come to the new city.[5] Factors will flow from the old to the new city until the two cities are of equal size, 1/2 $N(E)$ with the same K/N ratio.[6] (Note these factor flows are ensured because, in general, the rising parts of the factor reward paths are steeper than the declining parts, so that factor rewards in the initial city rise more slowly than those in the new city as factors flow from the initial to the second city.) Finally we note capital rentals and utility levels are both higher at 1/2 $N(E)$ than at $N(E)$.

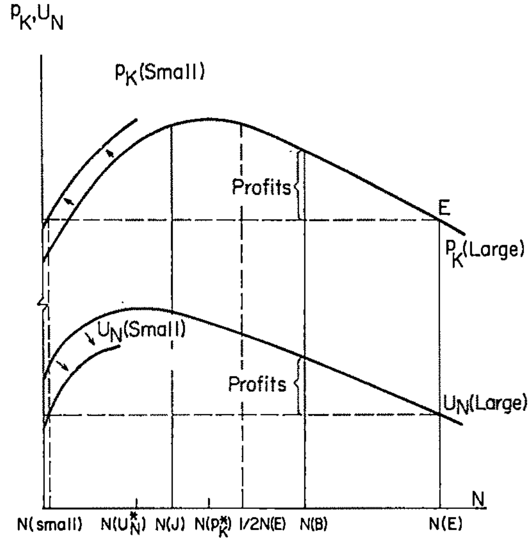


FIGURE 4. EQUILIBRIUM CITY SIZE

At 1/2 $N(E)$ the two cities continue to grow until they both reach size $N(E)$. At $N(E)$, by the above process, a third city forms. The resulting equilibrium has three cities of size 2/3 $N(E)$. As the economy grows new cities continue to form and the lower bound on *equilibrium city size approaches* $N(E)$, the point of city formation. Then for example, in the large sample case where the number of cities formally approaches infinity, *equilibrium* city size is at $N(E)$ in Figure 4. In contrast, under Assumption B, *optimum* city size lies between $N(U_N^*)$ and $N(p_K^*)$ at $N(J)$.[7]

Does divergence between equilibrium and optimum city sizes persist in a more sophisticated model?[8] The depiction of

[5] In a more sophisticated model there would be a speed of adjustment problem here. Suppose a firm does not instantaneously go out and form a second miniature city at point E in Figure 4. If our initial city size proceeds slightly beyond E, then two or more separately located small firm/cities become profitable at a point beyond E. This raises the possibility of three cities forming from the initial one. To avoid this problem, we assume that a firm acts as soon as the initial city reaches size E.

[6] Under Assumptions A and B, it is sometimes possible for factor rewards to be equalized with different K/N ratios in cities of the same type. For example, in Figure 3, our two cities could be of different sizes such that, with different K/N ratios and corresponding relative shifts in p_K and U_N, the curves p_K and U_N could be equalized between cities. In general, such solutions are ruled out as being unstable with respect to random factor movements.

[7] It has been pointed out several times to me that if the falling parts of the paths were very steep, market conditions would dictate cities splitting at a smaller size than pictured in Figure 4. This does not help the generalized rule for city formation since utility and capital rentals at $N(E)$ regardless of where $N(E)$ occurs are always the same as these factor records in an arbitrarily small city. Although the divergence from optimum city size might be smaller, the factor reward loss is just as bad.

[8] Note that our formation mechanism in terms of dynamics is naive, although it serves to reveal some of

firms or entrepreneurs acting myopically seems naive. Although scale economies are external to the firm, certainly there are entrepreneurs who will grasp the concepts of agglomeration benefits and disbenefits and be willing to initiate cities by moving industries, not just an arbitrarily small firm to form a new city. To facilitate our discussion, we introduce the "city corporation." The purpose of this exercise is to show there may be market forces ensuring that an equilibrium such as $N(E)$ would not persist. It does not pretend to deal with the dynamics of city formation!

D. *The City Corporation*

Suppose we are at $N(E)$ in Figure 4 in the large sample case. If a city corporation were to hire factors into a city restricted to size less than $N(E)$, factor rewards that could be paid in that city would be higher than competitive factor rewards. For example, at size $N(B)$, the corporation could hire factors competitively and have left over as profit the amounts indicated in Figure 4. Other entrepreneurs would follow the initial one, hiring factors and setting up new cities. Competition between entrepreneurs for factors to set up new cities should drive up factor prices and eliminate profits in the city corporation industry. Until cities are of size $N(J)$ where total benefits are maximized in Figure 3, by definition of $N(J)$, profits can be made by restricting city size. In other words, *the city corporation industry works "as if" the compensation mechanism used in the discussion of optimum city size is in effect.* If our city corporation industry is competitive and has adequate information, we will approach optimum city size $N(J)$. Note, however, to achieve this solution, the city corporation must be able to restrict city size. In the real world, either the corporation must own all land in the city or it must control land development and usage.

It seems likely that land developers play a crucial role in the real world and in a more sophisticated model than ours would play a more intricate role. For example, if our model allowed for suburbs, land developers would form suburbs as our core type or Mills-type cities grew in size. Suburbs would allow for (a) the release of pressure to form a completely new city due to rising commuting costs and (b) a mechanism for a completely new (economic) city to form where the "suburb" or our new city would be economically independent of the old city of the type in our model, without suburbs. By economically independent, we mean there would be little cross-commuting between the core city and suburb and interdependence in input and output markets would be weakened.

While there may be market forces ensuring attainment of optimal city sizes under Assumption B, this is *not* true under Assumption A. Equilibrium city size under Assumption A is determined in much the same way as for B. Laborers invest their capital throughout the cities of the economy so as to maximize capital rentals. On the other hand, they locate so to maximize the sum of U_N and U_K where p_K is exogenous to the labor location decision. To solve for equilibrium city size, we therefore use the p_K path to depict investment forces at work and the vertical sum of the U_N and U_K paths to depict the location or labor migration forces at work. We assume the existence of a city corporation mechanism and confine our discussion to the large sample case.

Parallel to the situation under Assumption B, with a city corporation mechanism, equilibrium city size will lie beyond

---

the forces at work. For example, technological change in industrial plants and transportation systems may be continually increasing optimum city sizes. Even if market forces leave us city sizes at $N(E)$, dynamic elements keep shifting optimum city sizes $N(J)$ out towards $N(E)$.

$N(U_N^*, U_K^*)$ at point $N(J)$ in Figure 5 where the decline in total utility or the sum of the decline in U_N plus U_K converted to dollar units becomes greater than the rise in $p_K\ \overline{K/N}$. The rise in p_K is given by equation (12). The decline in total utility equals $\partial U_N/\partial N$ plus $\partial U_K/\partial_N$ where p_K is fixed in the latter derivative, since laborers view $\overline{p_K}$ as exogenous to their location decision. The expression for $\partial U_N/\partial N$ is given by equation (14), while

$$\partial U_K/\partial N = a^a b^b c^c q_1^{-a} q_2^{-b}\ \overline{p_K}\ \overline{K/N}\partial(q_3^{-c})/\partial N$$

where $\overline{p_K}$ is given in (8); $\partial q_3/\partial N$ may be obtained from footnote 3, and $\overline{K/N}$ is the fixed ownership ratio of capital to labor. In making location decisions, laborers seek to maximize deflated real income or U_N plus U_K. In making investment decisions, laborers seek to maximize p_K. In trading off these two decisions or forces, city corporations maximize profits when the losses in location income of increasing city size are no longer exceeded by the gains in investment income.[9] Therefore equilibrium city size occurs at $N(J)$ where parallel to equation (16), we have

$$(18) \quad a^{-a} b^{-b} c^{-c} q_1^a q_2^b q_3^c \left| (\partial U_N/\partial N + \partial U_K/\partial K) \right| > \left| \partial p_K/\partial N \cdot \overline{K/N} \right|$$

The city corporation mechanism fails to solve the problem of laborers in their role as investors investing to increase city size to maximize capital rentals. This investment behavior inadvertently prohibits the attainment of optimum city size at $N(U_N^*, U_K^*)$. It is worth examining this result from another angle.

[9] Note that in the interests of profit maximization, city corporations could internalize the "externality" that occurs when individuals make location decisions with p_K viewed as being exogenous. In that case, $\partial U_K/\partial N$ in (18) would simply be the expression in equation (13). This of course does not help solve the problem that by investing to maximize capital rentals, laborers inadvertantly create a nonoptimal city size. It just alters the quantitative solution.

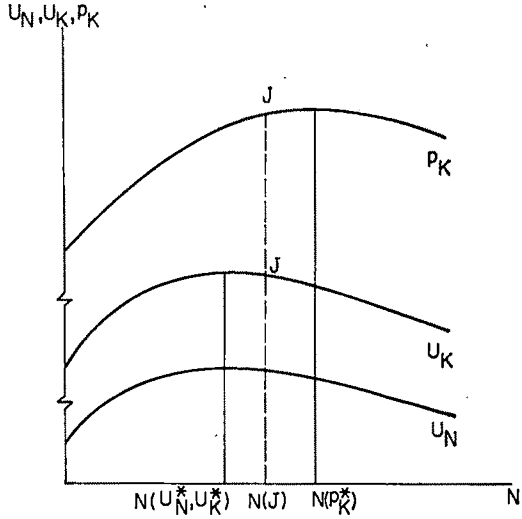


FIGURE 5. EQUILIBRIUM CITY SIZE: ASSUMPTION A

Point $N(J)$ is the equilibrium city size because no new profits can be made by entrepreneurs forming cities of a size different than $N(J)$ such as $N(U_N^*, U_K^*)$. For example, if a city corporation formed a city of size $N(U_N^*, U_K^*)$, it would raise the utility level of laborers *living* in the city. However, the capital rentals the city corporation could pay out would simultaneously fall. All investors could earn higher capital rentals which they are seeking to maximize in cities of size $N(J)$ and hence would not invest in a city of size $N(U_N^*, U_K^*)$. Similarly city size would not be bigger than $N(J)$ since the utility levels the city corporation could pay out would fall. As under Assumption B, by definition of $N(J)$ or the total benefit curve in Figure 3, the "sum" of utility levels from location decisions and capital rentals from investment decisions are maximized at $N(J)$ and no further profits can be made by changing city size.

IV. Extensions

In this section, we briefly outline complications that arise when we introduce a second (or more) type of city into the

model. Other extensions, not discussed here, include introduction of natural resources and transport costs of intercity trade.

Our second type of city specializes in the production of another type of traded good, say, X_2. The development of the utility and capital rental paths is the same as before. Equilibrium in the economy is depicted by both equilibrium in factor markets with equalized capital rentals and utility levels and equilibrium in output markets where markets clear and trade is balanced between the two types of cities. Different types of cities differ in size because production parameters, in particular α_i and ρ_i, differ between the traded goods of each type of city. Therefore, the shapes of factor reward paths determining city size will be different. Although utility levels will be equalized between cities, wage rates and housing prices will vary with city type and size.

Minor complications arise in the discussion of city formation.[10] When a new city of a particular type forms, factors from both types of cities will flow to it, since factor rewards will be affected throughout the economy. Equations (14)–(16) in this paper would have to be appropriately adjusted.

Other complications arise under Assumption A because the cost of living varies between types of cities. In equilibrium, capital rentals are equalized between all cities by investment behavior. Given these two facts, people living in larger cities will demand higher wages, not only because wage income is deflated by higher costs of living relative to smaller cities, but capital rentals are also deflated by higher costs of living. If we then allow capital owners still working as laborers to own different amounts of capital, there arises an incentive for laborers with relatively large dividend income to live in smaller cities or towns to enjoy the lower cost of living. In some sense, our distinction between Assumption B where capital owners live in the countryside and Assumption A may become rather fuzzy.

[10] Another complication arises when the economy is very small. The thrust of the argument is contained in our discussion of stability. When the economy is very small, it *may* be unstable to have in the economy two different types of cities both on the rising part of their factor reward paths.

REFERENCES

M. Beckmann, *Location Theory*, New York 1968.

G. H. Borts and J. L. Stein, *Economic Growth in a Free Market*, New York 1964.

J. Buchanan and C. Goetz, "Efficiency Limits of Fiscal Mobility," *J. Publ. Econ.*, Spring 1972, *1*, 25–45.

J. S. Chipman, "External Economies of Scale and Competitive Equilibrium," *Quart. J. Econ.*, Aug. 1970, *84*, 347–85.

J. V. Henderson, "The Types and Sizes of Cities: A General Equilibrium Model," unpublished doctoral dissertation, Univ. Chicago 1972.

W. Isard, *Location and Space Economy*, Cambridge 1960.

A. Lösch, *The Economics of Location*, New Haven 1954.

J. R. Melvin, "Increasing Returns to Scale as a Determinant of Trade," *Can. J. Econ.*, Aug. 1969, *2*, 124–31.

E. S. Mills, "An Aggregative Model of Resource Allocation in a Metropolitan Area," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 197–210.

R. A. Mundell, *International Economics*, New York 1968.

R. Muth, *Cities and Housing*, Chicago 1969.

C. Tiebout, "A Pure Theory of Local Expenditures," *J. Polit. Econ.*, Oct. 1956, *64*, 416–24.

R. Vernon, "International Investment and International Trade in the Product Cycle," *Quart. J. Econ.*, May 1966, *80*, 190–207.

A. Weber, *On the Location of Industry*, Chicago 1929.

Quality Competition, Industry Equilibrium, and Efficiency in the Price-Constrained Airline Market

By George W. Douglas and James C. Miller III*

Recent literature on regulated markets has focused on the ramifications of controlling price as a means of regulating a monopolist's rate of return.[1] There exist, however, a great variety of regulated *competitive* markets where the implications of price control are qualitatively different. That is, while the control of price to restrain a monopolist's profit rate may affect the technical efficiency of its production of a specifically defined output, the implications of price restraints in nonmonopoly markets are more typically manifested in the overall *quality* level of the output, an issue of allocative efficiency. There are many examples of such regulated markets (and imperfect private cartels) which, although price constrained, reach an equilibrium through vigorous nonprice competition which bids away most or all implicit rents. In these markets the existence of rents (or the regulator's control of rents) is more directly related to the nature of restraints on entry rather than to the control of price. Examples include commercial banking, stock brokerage, real estate brokerage, motor trucking, taxicab service, and, of particular interest here, air transportation.

By describing industry equilibrium in these nonmonopoly markets one can identify an endogenous relationship between the regulated price and the overall general quality level of the output. For many of these markets a viable, zero-rent equilibrium can obtain over a wide range of prices, each defining implicitly an endogenous quality level. Hence, regulators indirectly control quality by the selection of the price parameter. Neither efficient costs nor optimal prices can be defined without reference to this price-quality relationship. While regulators may attempt to control tangible aspects of output quality by direct fiat, the endogenous quality determined in equilibrium is not usually amenable to direct control, if it is perceived at all. Complicating this process is the difficulty of finding an appropriate, measurable proxy for the quality dimension involved.

Equilibrium in the domestic airline markets under Civil Aeronautics Board (*CAB* or Board) regulation typifies this phenomenon and illustrates the rather wide divergence between actual and efficient prices and quality levels that may obtain under regulation. Although entry restraints have insulated the trunk carriers from new firm competition and at present the typical city-pair market is served by only a relatively small number of firms, vigorous nonprice competition has prevented the carriers from obtaining the rents which are potentially available in the regulated price structure. On the one hand, the carriers compete in the traditional, highly visible forms of nonprice rivalry such as advertising and the provision of passenger amenities. But of far greater sig-

* Associate professors of economics, University of North Carolina at Chapel Hill and Texas A&M University, respectively. We gratefully acknowledge helpful comments from John Evans, an anonymous referee, and others. A portion of this research was supported by the Brookings Institution.

[1] See the concise literature summary contained in William Baumol and Alvin Klevorick.

nificance in terms of per passenger cost is their use of scheduling competition in establishing market shares. While analytically similar to advertising competition, scheduling competition has a fundamental role in determining a basic characteristic of service quality in the scheduled system.[2]

By its very nature, scheduled transportation is nearly always characterized by some excess supply, or slack capacity. Slack capacity arises from several sources. Supply (capacity) is produced in discrete units and cannot be stored, while demand is a flow which fluctuates seasonally and stochastically. Moreover, as supply is augmented on any given route, service convenience is enhanced by the more frequent departures and the greater likelihood of the potential passenger's obtaining a seat on the preferred flight. The significance of scheduling competition among the carriers then is its effect on the system's overall level of service convenience. Since the relative slack capacity associated with equilibrium is related to the regulated price, the description of airline costs, and the selection of an efficient price level and structure require an assessment of the relationships among slack capacity, passenger costs, and overall service quality.[3] In this paper we describe and attempt to quantify these relationships, suggest the optimal level and structure of air fares, and contrast this estimated efficient solution with the market equilibrium now obtaining under regulation.

I. Quality Competition and Airline Market Equilibrium

A. *The Demand for Airline Services*

The demand for air travel in any market (i.e., point A to point B) can be regarded as a flow over time which fluctuates seasonally (given the time structure of prices) and stochastically. Thus, the time of consumption is a significant dimension in the definition of scheduled travel output. The expected volume of demand over any reference time period depends, as usual, on the price (fare) and the tangible aspects of quality, such as (expected) en route time, comfort, etc. But while demand arises continuously over time, any scheduled system provides capacity at discrete intervals in lumpy units; hence, most travellers must depart at a time other than that which is the most preferred.[4] The degree to which these discrete units of supply are matched with the continuous demand may be regarded as the "convenience" of the scheduled system. Such convenience, of course, is a highly significant determinant of total airline demand, particularly in markets where reasonably close substitutes exist. Convenience, moreover, can be measured. We shall define *schedule delay* as the absolute difference between the passenger's most preferred departure time and that of his actual departure. (Thus our inconvenience measure accounts for passengers who find they must revise their preferred plans and choose either to advance their departure or delay it.)[5] Then

[2] The role of scheduling competition in determining equilibrium in this market has been noted by various writers, including Arthur De Vany (1968), George Eads, Miller (1969), Lawrence White, and Joseph Yance (1972). In this paper we formalize the relationship and develop the quantitative implications for service quality arising from the regulated fare.

[3] The recent airline cost study by Theodore Keeler is a case in point. Since the cost per passenger carried varies markedly with the average load factor realized, the estimates of "efficient," cost-based fares depend crucially on the hypothesized load factor of an efficient system. We indicate in this paper why it is inappropriate to extrapolate the quasi-competitive load factors of the California intrastate markets as an efficient norm for all markets.

[4] Herein lies the *raison d'être* for the institution of scheduled transportation: while demand for travel originates continuously over time and space, aggregation of demand over these dimensions is necessary to exploit the considerable economies of transport with respect to vehicle size and utilization.

[5] Since a priori we do not know whether travellers are symmetric in their time preferences (i.e., whether on balance they view equally an adjustment to an earlier

for any equilibrium of a scheduled system, there will exist a distribution of schedule delays. An expected value of this distribution—expected schedule delay per passenger—serves as a proxy for the overall level of service convenience.[6]

Let us denote the demand for scheduled air services in a specific market as:

$$N = N(P, T, X) \tag{1}$$

where

N = number of air passenger trips demanded per time period
P = average fare
T = expected schedule delay per passenger (i.e., (in)convenience)
X = vector of tangible aspects of service quality

Note, however, that expected schedule delay is not an exogenous variable controlled directly by any firm (in nonmonopoly markets), but is an outcome of the market equilibrium. That is, schedule delay is related to the total frequency of departures in the market and to their timing. Moreover, it is also related to the probability of being unable to obtain a seat on a scheduled flight. Since demand is stochastic, we observe a typical queuing phenomenon, in which the probability of obtaining accommodation on the convenient departure flight depends on the market equilibrium's level of slack capacity, the dispersion of the demand, and the schedule frequency.[7]

Although we will describe the process more completely below, we can express the implied delay relationship generally as

$$T = T(N, \sigma_N, F, S) \tag{2}$$

where

T = expected schedule delay per passenger (for example, in minutes)
N = number of passenger-trips demanded per time period
σ_N = dispersion of demand
F = total flight frequency
S = average aircraft capacity

A priori, expected schedule delay per passenger would have the following qualitative relationships with the independent variables:

$$\partial T/\partial N > 0, \qquad \partial T/\partial \sigma_N > 0,$$
$$\partial T/\partial F < 0, \qquad \partial T/\partial S < 0$$

We might suppose that the stochastic nature of demand generation for a given market would allow us to express as a function:

$$\sigma_N = \sigma_N(N) \tag{3}$$

By substituting equations (2) and (3) into the demand equation (1), we obtain the following demand expression:

$$N = G(P, F, S, X) \tag{4}$$

B. *The Costs of Producing Airline Services*

For practical purposes, airline total costs can be dichotomized into the cost of providing capacity on a route, and the cost of providing passenger services. (See Miller (1972), pp. 224–28, and our 1974 paper, ch. 2.) Firm i costs may be written:[8]

or later departure time), we shall weigh them equally per time period expended.

[6] We note that schedule delay is not the only type of delay impacting on overall service quality. Certainly delays in departure, in receiving on-board services, in arrival, and in baggage claim are important. In this analysis we treat these delays as exogenous. Since reductions in such delays (or their expected value) are a means of nonprice competition, their overall levels could be treated in a way similar to the scheduling competition discussed here.

[7] We do not infer that travellers actually form a queue in the terminal. The phenomenon is similar, however, even with the existing reservation process. Stochastic fluctuations result in excess demand for some flights, and some travellers must displace their departure time from the most preferred. Alternatively, we could measure the phenomenon as the advance time required to obtain a reservation with some arbitrarily specified level of certainty.

[8] Note that we have not made X a firm variable. There are several reasons why we might make this sim-

$$(5) \qquad C_i = C_i(N_i, F_i, S_i, X) = f_i(F_i, S_i) + g_i(N_i, X)$$

To analyze equilibrium under nonprice (scheduling) competition, it is convenient to assume that for a specific market the selection of aircraft size is exogenous.[9] Moreover, a number of cross-section studies of airline costs suggest that there are no scale economies associated with firm size.[10] This allows us to rewrite cost function (5) as

$$(6) \qquad C_i = cN_i + kF_i$$

where c and k are both constants.

C. *The Nature of Regulatory Restraints*

Although the Board has never allowed the entry of a new trunk carrier, it has allowed the entry of several classes of carriers which now (to varying extent) compete with the trunks (for example, the local service carriers). More importantly for our purposes here, the Board has certificated entry of existing trunks into individual city-pair competition to the extent that presently less than 25 percent of trunk carriers' traffic is generated in quasi-monopoly markets. (See exhibits by *CAB* Bureau of Operating Rights.)

CAB control over price, however, is so pervasive that carriers tend to view price as a parameter and compete (or rival) almost entirely in nonprice dimensions, primary of which is the frequency of scheduling. (See our 1974 paper, ch. 6.) It is notable that while on various occasions the Board has limited the extent of carrier competition over passenger amenities, the Board is prohibited by law from regulating (directly) the frequency of carrier scheduling.[11]

D. *A Model of Firm Behavior and Industry Equilibrium*

With price P^* as a parameter, the first-order condition for firm i profit maximization can be written

$$(7) \qquad \partial\Pi_i/\partial F_i = (P^* - c)\partial N_i/\partial F_i - k = 0$$

Simply stated, the marginal revenue from an additional flight $(P^*-c)\partial N_i/\partial F_i$ must equal the marginal cost k. Alternatively, profit maximization may be interpreted as equating the exogenous price P^* with the marginal cost of *attracting and carrying* an additional passenger $c+k/(\partial N_i/\partial F_i)$. Profits exist if the marginal cost as perceived by the firm exceeds the average cost per passenger $c+k(F_i/N_i)$.[12] This occurs if the perceived effect on the firm's demand of the marginal flight $\partial N_i/\partial F_i$ is less than the average number of passengers per flight N_i/F_i. For the monopoly market, we would expect this condition to hold in equilibrium, since at some point (if not from

plifying assumption. First, some dimensions of service-amenity competition come under direct *CAB* regulation (for example, "free" drinks). Second, there exists little real difference among carriers regarding the passenger amenities offered. Finally (and in partial explanation of the second reason), one observes typically that the extent of passenger service amenities varies closely with the extent of scheduling competition (i.e., both are means of attracting passengers). In short, it would appear that in price-restrained "competitive" markets, firms are "quality takers."

[9] While aircraft size is endogenous to the overall equilibrium, the interdependence of production across markets served would tend to diminish the influence of specific markets. Rather, carrier fleets typically contain a small number of aircraft types, each of which dominates in certain classes of markets.

[10] Studies by Richard Caves; Eads, Marc Nerlove, and William Raduchel; Mahlon Straszheim; and Miller (1970) indicate that average costs of the airline firms are not significantly related to the firm's level of aggregate output. One cannot infer from these studies, however, that scale economies do not exist in specific city-pair markets.

[11] Section 401(e)(4) of the Federal Aviation Act of 1958 (as amended) prohibits the Board from regulating carrier schedules. However, under Section 412 of the Act the Board may approve industry agreements to "self-regulate" (for example, the airport congestion agreements and the recent agreements whereby three trunk carriers have reduced capacity in four transcontinental markets).

[12] This follows the analysis of nonprice competition by George Stigler.

the beginning) there would be diminishing returns to scheduling: $\partial^2 N/\partial F^2 < 0$.[13]

The extent to which scheduling competition bids away implicit rents, therefore, is determined crucially by the carriers' perception of the impact on firm demand of a unilateral increase in flight frequency. If the individual firm in a multifirm market believes that its schedule change will not be met by its rivals, then its perceived marginal cost will be less than that of a monopolist, since the effect on the firm's demand includes "diversion" from its competitors, as well as generation of new travellers.

To see this, let firm i's demand be

$$N_i = \lambda_i N(P, T, X) \tag{8}$$

where T is defined by (2) and where λ_i, the market share of firm i, is given by some function

$$\lambda_i = \lambda(F_i/F, F_j/F, \ldots, F_n/F) \tag{9}$$

Then

$$\partial N_i/\partial F_i = \lambda_i(\partial N/\partial F) + (\partial \lambda_i/\partial F_i)N \tag{10}$$

If, for example, the firm's market share were equal to its share of the total flights scheduled in the market, equation (10) becomes

$$\partial N_i/\partial F_i = (F_i/F)(\partial N/\partial F) + (1/F - F_i/F^2)N \tag{11}$$

which may be rewritten

$$\partial N_i/\partial F_i = \{F_i/F + [1 - F_i/F] \cdot [(N/F)/(\partial N/\partial F)]\}\partial N/\partial F \tag{12}$$

Since at equilibrium $N/F > \partial N/\partial F$, the expression within the braces of (12) exceeds unity. Hence, the perceived impact of a unilateral increase in frequency by a competitor (not met by its rivals) exceeds that of the monopolist. Moreover, this perceived impact increases as the number of competitors grows (i.e., as F_i/F diminishes). Hence, the level of total market frequency at which for all firms the perceived (marginal) revenue of an additional flight equals the flight's marginal cost grows with the number of competitors.[14]

This bias in perceived effects of scheduling is greatly amplified, moreover, by a carrier management-perceived non-linear market share function. Cross-section observations of market shares in competitive airline markets would seem to imply that a carrier with the largest share of flights attracts more than a proportional share of the traffic.[15] To test this notion, cross-section observations of 137 city-pair markets in 1969 were fitted to the following market share function:[16]

$$\lambda_i = N_i/\sum_{j=1}^{n} N_j = F_i^{\alpha}/\sum_{j=1}^{n} F_j^{\alpha} \tag{13}$$

thus yielding an estimate $\alpha = 1.233$, with a coefficient standard error of .018 and an equation $R^2 = .959$ (with 401 degrees of freedom). The estimate of α significantly exceeds unity, the value it would take if the market share relationship were proportional.

This market share-frequency share result has an important implication for airline scheduling strategy, for it increases a carrier's incentive to add schedules and as a result compete away industry rents. Whatever the relative market shares, as long as industry rents are being earned one or more carriers in a multifirm market can raise its own load factor (and increase its profits) by expanding service. The carrier

[13] Competitive market equilibrium likewise requires that $\partial N/\partial F < N/F$, else scheduling would increase without limit.

[14] Note, however, that if each firm believes that any change in flight schedules will be met by its rivals, the monopoly solution obtains.

[15] This phenomenon has been described previously in cross-section studies by Gilles Renard, among others.

[16] This specification is that of Renard. The data were taken from carrier reports of 1969 load factors by city-pair market as summarized in an exhibit by the *CAB* Bureau of Economics (1970b).

which fails to be aggressive when others are finds its market share significantly diminished. This incentive for unilateral capacity augmentation makes collusion difficult to orchestrate and enforce.[17] Since $\partial N/\partial F < N/F$, capacity expansion lowers market average load factor and eliminates rents. Thus, to the extent this non-linear market share-frequency share relationship is perceived by the carriers, the achievement of zero-rent equilibrium would require the existence of only a small number of competitors.[18]

E. *Evidence on the Model*

Observations of market data are consistent with the above hypothesis of market equilibrium. First, over time the trunk carriers have bid away potential rents, as indicated by their reported profit rates. For example, over the period 1955 through 1970, the trunk air carriers averaged only 6.42 percent return on investment (i.e., net revenue after taxes but before interest payment on debt, as a proportion of equity-plus-debt). (See *CAB* (1972, p. 76).)

Second, cross-section observations of city-pair markets reveals a pattern consistent with the zero-rent hypothesis. If each specific market were independent (i.e., could be treated as a unit separate from the airlines' system of routes), we would observe in each an average load factor that approaches the break-even load factor for that market (predicted by the average fare and estimated production costs, including a normal profit). Although most airline markets are not completely independent (especially in the production of capacity), we do observe patterns in the cross-section equilibria that are consistent with those hypothesized. For example, while fares per passenger mile are lower on long hauls than on short hauls, they have not fully reflected the economies of production with respect to distance. Hence, the break-even load factors are lower for long hauls than for short hauls. As a result of intensive scheduling competition, however, actual load factors are bid down on the longer and denser routes which otherwise present the potential for the greatest rents.[19]

Third, average load factors (and rents) tend to be inversely related to the number of firms in the market. The cross-section regressions in Table 1 illustrate the effects

TABLE 1—CROSS-SECTION ANALYSIS OF MARKET AVERAGE LOAD FACTORS

(t-statistics in parentheses)

$$1)\ ALF = .588 - \underset{(1.4)}{2.11 \times 10^{-5} D} + \underset{(9.1)}{7.62 \times 10^{-7} N} - \underset{(6.5)}{7.06 \times 11^{-2} C}$$
$$R^2 = .213 \qquad Df = 347$$

$$2)\ ALF = .257 - \underset{(1.8)}{.019 \ln D} + \underset{(7.1)}{.073 \ln N} - \underset{(5.5)}{1.46 \ln C}$$
$$R^2 = .144 \qquad Df = 347$$

Source: CAB, Bureau of Economics (1970a).
ALF = market average load factor (i.e., total passengers/total seats)
D = market distance (in miles)
N = average number of daily passengers in market
C = number of carriers in market

[17] For example, as mentioned by Eads, when in 1971 the Board granted antitrust immunity to three trunk carriers for the purpose of drawing up capacity limitation agreements in fifteen markets, agreements could be reached on only four.

[18] A vice-president of TWA, Melvin Brenner, acknowledged management's belief in this S-curve market share phenomenon in testimony before the *CAB*. He suggested, moreover, that this phenomenon generates loss equilibria in many markets in a "prisoner's dilemma" context. (See also William Fruhan, ch. 5.) This has been shown by Yance (1970) to be possible, but only in small markets with an improbably large number of competitors.

[19] The industry's fare structure originated from the pattern of first class rail fares prevailing thirty years ago. (See Caves.) Changes in fares have increased the "taper" of fares with distance, yet the Board's staff tends to perpetuate the phenomenon by computing average costs on the basis of previously realized load factors.

of distance, density, and number of carriers on the markets' average load factors. (Note also in Figure 3 the hypothesized effects of distance and number of carriers.)

Finally, we do observe from time-series that actual load factors tend to equal and follow changes in the estimated breakeven levels. (See Air Transport Association, Slide 16.) Where substantial divergences do exist they tend to be explainable in terms of lags in equipment acquisition (or disposition) or differences between *ex ante* and *ex post* demand.

To recapitulate, in a market in which scheduling competition bids away all rents, the regulator in choosing price implicitly determines the equilibrium number of travellers and the expected per passenger schedule delay. The regulator's role, therefore, is one of serving as a proxy for the population of consumers in choosing the appropriate combination of service quality and price from the opportunity locus of these variables.[20] We consider below and attempt to quantify the opportunity locus between expected per passenger schedule delay and price.

II. Determining an Efficient Price-Quality Configuration

A. *Schedule Delays and Their Cost*

In order to obtain the tradeoff between equilibrium price and expected schedule delay, one must first approximate delay function (2). We analyze schedule delay as arising from two sources: the difference between a traveller's desired departure time and the closest scheduled departure, and the delays caused by excess demand for the flight(s).

The former delay, denoted as *frequency delay*, we have estimated by simulation. To do this, we superimposed an assumed scheduling of F departures on a representative time pattern of demand.[21] The number of flights scheduled in the simulation was varied from unity to 50 daily. For each value of F the average *absolute* differential between the traveller's preferred departure time and the closest scheduled departure was computed as an estimate of expected per passenger frequency delay, $T(F)$. The estimated values of expected frequency delay were then fitted to the daily frequency, F, from which they were generated:[22]

$$(14) \qquad T_f = 92(F^{-.456}); \qquad R^2 = .497$$
$$\text{(standard error of exponent} = .119)$$

The variable T_f is the expected absolute value of the time differential in minutes between a scheduled flight and the traveller's preferred time of departure.[23]

Because demand fluctuates stochastically, in some instances a specific flight

[20] Insofar as quality differentiation is constrained, especially with regard to schedule delay, the regulator is obliged to choose a single common quality level for a population whose preferences for quality may be diverse. This issue is discussed by Douglas (1972) with respect to a taxicab market and is raised again in the conclusion of this paper.

[21] The time pattern employed may be found in Miller (1972, p. 233). Ideally, we wish to identify the pattern of expected flows by time of day that would be observed if there were a continuous stream of departures with excess capacity in each. Traffic data, of course, are influenced by the pattern of departures as well as demand. The time pattern used for this exercise is heuristic, reflecting, however, typical market data. (See traffic data by the Department of Transportation.)

[22] The F flights were scheduled such that each flight faced an equal number of potential travellers. Alternatively, the flight assignment could have been made to minimize expected delays. In a real market, however, the assignment of flights would reflect the time pattern of fares and the cycling time of the system (or round trip time in a two-node system). It might be noted further that while the expected delay with a uniform time distribution of demand would be linearly related to the inverse of the flight frequency, in a market where the time profile of demand is "peaked," the relationship is not necessarily linear. Moreover, peaking of demands reduces the expected frequency delay.

[23] Alternative scheduling practices may make the expected delay larger or smaller. However, the importance of the delay function is in estimating *changes* in delay as daily frequency is changed (i.e., $\partial T_f/\partial F$). While the absolute value of the estimate may be biased, its marginal change may not be.

faces excess demand. In such a case some travellers must look for the next best departure, and, if it is filled, the next best, etc. A simple queuing model was developed in which the process was characterized as a Markov chain. (See Douglas (1971, Appendix A) and the authors, Appendix 6-A.) From the Domestic Passenger Fare Investigation (*CAB* Docket 21866), data are available on daily passenger traffic in selected markets during February and November 1969. These data were found to be consistent with a Poisson process with the sample variance linearly related to the sample mean by

$$\text{var}(n) = 16.978\,\bar{n} \tag{15}$$

We could thus assume that the distribution of demand arising within any time interval τ would be Poisson, with

$$\sigma_n \simeq 4.12\,\bar{n}^{1/2} \tag{16}$$

This allows us to define a Markov process, in which the state of the system is described by Q, the number of (potential) passengers drawn from the demand distribution for that flight *plus* those passengers who were unable to obtain seats on previous flights. The steady state of the system estimates the probability distribution of the queue length facing a flight.

The probability of being delayed one flight interval, two flight intervals, etc. can thus be calculated.[24] Given the average time interval between flights, we can then estimate the expected displacement of a traveller's actual departure time from the closest scheduled departure. The absolute value of this displacement, which we denote as *stochastic delay*, was estimated over many system configurations of demand, demand dispersion, schedule frequency, and aircraft size. These estimates of stocastic delay were fitted to the function:

$$T_s = .455(Y^{-.645})(X^{-1.79})(I); \tag{17}$$

$$R^2 = .910, \qquad Df = 44$$

where

T_s = expected stochastic delay per passenger (in minutes)

$Y = N_f/\sigma_f$

$X = (S - N_f)/\sigma_f$

I = average interval between flights (in minutes)

N_f = mean demand per flight

σ_f = standard deviation of demand per flight

(Standard errors for the exponents of Y and X were .114 and .100, respectively.)

The sum of frequency delay T_f and stochastic delay T_s we denote as (total) expected schedule delay per passenger T:

$$T = T_f + T_s = 92(F^{-.456}) + .455(Y^{-.645})(X^{-1.79})(I) \tag{18}$$

In Figure 1 we indicate the estimated

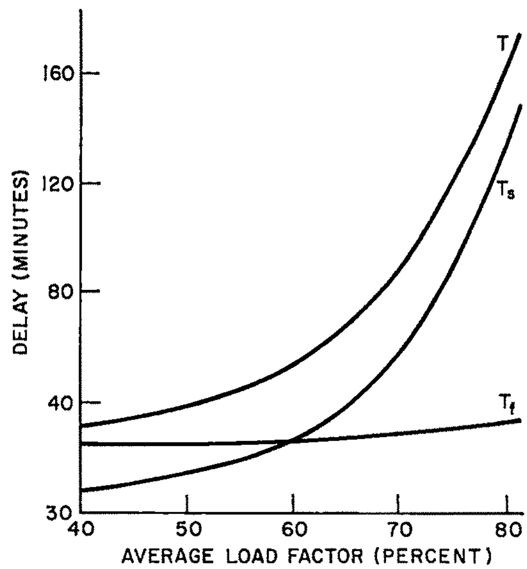


FIGURE 1

[24] Using the Markov chain we are implicitly assuming that travellers always delay their departure if a flight is filled. The estimate thus overstates the expected stochastic delay if one considers that some passengers may shift to earlier flights. This becomes, in effect, a conservative bias in the later estimate of "optimal" slack capacity.

values of the components of schedule delay as the excess capacity of the system changes in a hypothetical market.[25] We note particularly the asymptotic increase in stochastic delay T_s as the relative slack capacity becomes small (i.e., as average load factor approaches unity).[26]

B. *The Price-Quality Tradeoff*

In scheduled transportation, the average cost per passenger carried is quite sensitive to the utilization of capacity, or the relative level of slack capacity in the system. The relationship between average (per passenger) cost and average load factor is, of course, inverse: the higher the average load factor, the lower the per passenger cost.[27] By combining empirical estimates of the relationship between per passenger cost at various levels of slack capacity (i.e., average load factor) with the expected delay function we are able to estimate the opportunity locus between average (and marginal) cost (price) and expected schedule delay (quality).[28] The tradeoff for a hypothetical market is illustrated in Figure 2.[29]

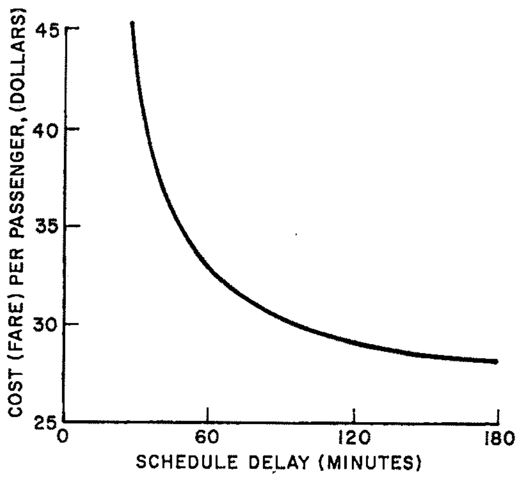


FIGURE 2

An efficient price-quality equilibrium would be characterized by the equality of the technical tradeoff between price and quality (schedule delay) in a market with the subjective tradeoff of the passengers.[30] The latter could be derived from an appropriately estimated demand function for air travel (1):

$$\rho = (\partial N/\partial T)/(\partial N/\partial P) \tag{19}$$

While the data required to fit the demand function specified are unavailable, we may learn much about the optimal price-quality configuration (and thus the relative efficiency of the existing market equilibrium) by assigning a reasonable value to schedule delays. In this way one can, at the very least, describe qualitatively the optimal levels of slack capacity

[25] In Figure 1, T is (total) schedule delay, T_s is stochastic delay, and T_f is frequency delay. This hypothetical market has a distance of 600 miles and serves 800 passengers per day with three-engine turbofan aircraft.

[26] Note also that T_f rises slightly with increasing load factor. The reason for this is that in the figure the number of daily passengers is held constant and thus to increase load factor schedule frequency must fall and average frequency delay rises.

[27] That is, $AC = c + k/L$ where L is average load factor (i.e., $L = N/Fs$ where s is the average number of seats per flight).

[28] There are several sources of airline cost functions suitable for estimating the average cost–average load factor tradeoff. (See *CAB*, Bureau of Economics (1970b).) For the analysis summarized in this paper we utilized the cost model developed in the authors, ch. 2. Consistent with a regulatory cost-recovery restraint, price must equal average cost. We note also that for a *given* load factor service $AC = MC = c + k/L$ where $L = N/Fs$.

[29] For any given market with an underlying time pattern of demand there exists a cost function, $C = C(N, T_e, X)$, which describes the minimum cost of transporting the demand flow N, with a level of expected delay T_e, and tangible quality X. Technical efficiency requires that the market equilibrium reach a point on this frontier. The opportunity locus described in Figure 2 may be interpreted as a two-dimensional representation of the average cost function implied. Figure 2 represents a hypothetical market with distance of 600 miles, serving 800 passengers per day with three-engine turbofan aircraft.

[30] If all passengers do not have the same subjective price-quality tradeoff, then a weighting problem exists. See discussion in Section V, regarding the desirability of increasing the number of price-quality options.

in markets of differing distances and density. Moreover, within certain boundaries, one can appraise the overall appropriateness of prices and quality that have obtained under regulation.

III. Efficient vs. Existing Price-Quality Level and Structure

Critics have often asserted that the airline industry is characterized by "excess" capacity. Clearly, such an assessment requires further analysis of the price-quality tradeoff described above. Casual observation of the interaction between the industry and the regulators, moreover, suggests why such a bias might exist. The market shares competition previously described will often generate short-run losses, especially when equipment acquisition is overoptimistic or when expected demand fails to materialize. If the Board attempts to prop up carrier profits with a general rate increase, it implicitly rationalizes the lower load factors and sets the stage for another iteration of higher fares and increasing capacity. A review of the Board's rulings on fare increases confirms that carrier profits are a principal target of policy. (See Caves and William Jordan.)

In Figure 3 we indicate the estimated optimal load factors for hypothetical markets of equal density but of variable distance. These are defined as the load factors which minimize the sum of average cost per passenger (fare) and the value of the expected schedule delay for each of two assumed values of schedule delays. In a market with average daily demand of 800, efficient load factors increase with distance, ranging in level from approximately 57–62 percent at short distances to approximately 69–72 percent in transcontinental markets. These estimates are predicated on an imputed passenger valuation of schedule delays of \$5 and \$10 per hour.[31]

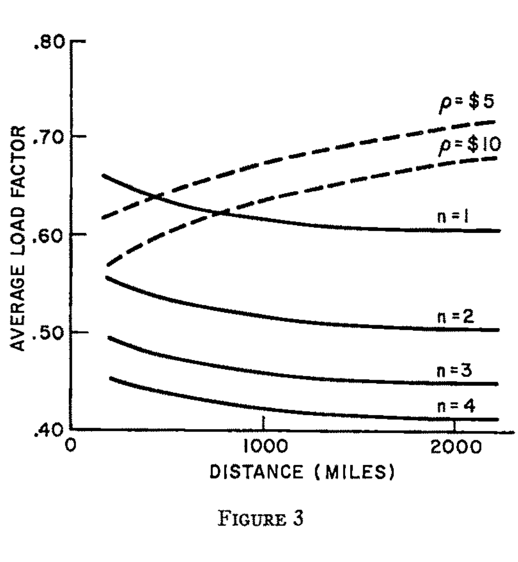


FIGURE 3

The rise in efficient load factors with distance may be explained by the following. Expected schedule delays are a function of the relative number of empty seats carried, on the average. Since the cost of carrying an additional empty seat on a 2000 mile flight exceeds that of a 500 mile flight, one chooses to buy more quality (i.e., carry more empty seats) on the latter than on the former. This pattern is to be contrasted with that obtaining under regulation indicated by the plotted regressions in Figure 3, derived from regression 2) in Table 1.[32]

[31] While we refer to ρ as the value attributed to schedule delay, we do not imply that it is equivalent to the traveller's value of en route time. The time displacement here is qualitatively different; we should expect however that having alternative uses, these delays would be valued at less than that of en route time. De Vany's 1968 measure of en route time value, \$7.28 (inflated to \$10 in round numbers for 1971), serves generally as an upper bound for schedule delay time value. See De Vany (1974).

[32] In Figure 3 the estimated optimal values (dashed lines) were computed with alternative valuations of schedule delay ρ, \$5 and \$10. Markets were presumed to have a demand density of 800 daily passengers; two-engine turbofan aircraft were assigned for distances under 1,000 miles, three-engine turbofans for distances 1,000–1,800 miles, and four-engine turbofans for distances in excess of 1,800 miles. In the figure, n denotes the number of carriers in the market. We might note that while the explanatory value of the regressions in Table 1 are low, when separate regressions are run on the data segmented by number of carriers in the market the correlation coefficients generally increase. In all

Optimal slack capacity also varies with demand density, or market size. That is, for a given level of relative slack capacity the probabilities of being delayed by j flight intervals is independent of market size, yet the interval between flights diminishes with market density. Hence, in larger markets one can obtain more quality with relatively fewer empty seats, and the optimal load factor increases with market size. This relationship for hypothetical markets is illustrated in Figure 4.[33] Note the contrast between optimal load factors and those existing under regulation (as derived from regression 2) in Table 1).

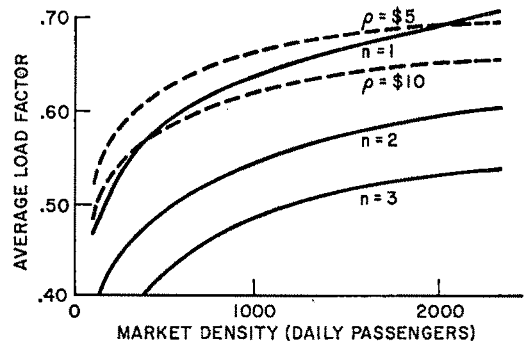


FIGURE 4

The contrast between the regulated equilibrium and the efficient equilibrium is dramatically demonstrated by the intrastate California markets. A price competitive equilibrium has emerged in some of these dense markets with the activity of a maverick firm, Pacific Southwest Airlines. The prices in these markets have ranged from 40 to 50 percent less than in similar markets under *CAB* regulation. This difference can be attributed principally to the higher load factors in the California markets (in the 65–75 percent range), which is consistent with our estimate of an efficient price-quality combination for markets of this character.[34]

Implicit in this description of scheduled transport markets is an aggregate cost of service function. Technical inefficiency in the provision of service can arise from numerous sources, including inefficient timing of schedules (for example, Hotelling-type clustering), inappropriate selection of fleet composition, or inefficient routings. While the potential exists for technical inefficiency in all these forms under regulation, it is notable that an unregulated airline market might not achieve technical efficiency either.[35]

Another implication of the model is that "social" scale economies exist in the production of scheduled air service even if there are no private scale economies accruing to the airline firm. That is, the average cost per passenger carried with a given level of schedule delay T and tangible quality X decreases as market size increases. Using the methodology outlined earlier we can derive a relationship between average and marginal costs of carrying an additional passenger, *while holding expected schedule delays constant.*[36]

cases except four-carrier markets there is an inverse relationship between average load factor and distance. (See the authors, ch. 4.) Even in these markets, however, the overall levels of average load factor are much lower than reasonable estimates of the optimal levels.

[33] In Figure 4, optimal average load factors (dashed lines) were estimated on the basis of two valuations of schedule delay, $\rho=\$5$ and $\rho=\$10$. Market distance is 600 miles and service is provided by two-engine turbofan jet aircraft. In the curves tracing out actual average load factors, n denotes the number of firms in the market.

[34] The California experience is described at length by Jordan. He fails, however, to relate the fare dichotomy to excess capacity and service quality. Aside from the question of X efficiency of the firms sheltered by regulation, similar results could obtain under *CAB* regulation. The low fare and high utilization which characterizes the markets to Puerto Rico is such an example.

[35] While the observed competitive markets do appear to yield a more appropriate balance of price and quality, it is not clear but that the nonproportional market share function would bias the fleet selection and scheduling practices of carriers in many nonregulated, competitive markets. See Douglas (1972) for a discussion of the competitive equilibrium likely to be found in a competitive transport market.

[36] Herbert Mohring's description of the costs of a scheduled bus system arrives at a similar conclusion. His analysis, however, does not consider the stochastic component of delays.

The results imply that the marginal (social) cost of a given service quality is, on average, some 10 percent less than average cost. In principle, this might suggest a rationale for the (third-degree price discriminating) discount fares now prevalent in the industry, under the assumption that the industry be operated with a cost-recovering revenue restraint. However, the implied discounts are certainly less than have existed, and in any event external costs of production might easily vitiate this result.

IV. Conclusion

A description of efficiency in price-constrained "competitive" markets requires an understanding and assessment of the endogenous relationship between quality and price. In the scheduled air transport industry, an important component of quality is endogenous to the equilibrium, and a proxy measure of this quality characteristic can be estimated. We note, however, that our estimates of the optimal price-quality configuration and thus comparisons with the existing regulated-market equilibria are based on information a good deal less complete than ideal. Specifically, we would prefer to have additional information on the value passengers place on delay time and more complete data on demand distributions by market characteristic. What we have attempted to show is that ignoring the price-quality tradeoff can lead (and probably has led) to significant divergencies between the optimal and existing price-quality options. With more complete information one could be more precise in such an efficiency assessment.

One implication of this analysis is that, *ceteris paribus*, the greater the number of price-quality options the better. We note that presently the market does provide air transportation in various configurations, characterized by wide differences in cost and schedule delay. At one end of the scale, charter flights provide the lowest price, but the greatest schedule delay. At the other extreme, corporate executives are whisked about the country in private jet aircraft at enormous cost, but with a very minimum of schedule delay. Scheduled air transportation lies between these poles, and, as we have suggested here, at the margin of excessive quality and price.

Some quality differentiation, moreover, does exist within the scheduled transport industry in this dimension. Stand-by fares and "leisure class" travel represent ways in which differentiation with regard to stochastic delays is currently practiced, although in a very limited way. Of course, additional techniques could be developed for price-quality differentiation within the regulatory environment, and in general such initiatives are to be encouraged.

Finally, it is clear that the costs of the regulated carriers are high *because* the price level is high, and not the reverse (i.e., cost is price-determined, not price-determining). Further, the proclivity of the carriers to compete intensively with scheduling rivalry implies that the regulator need not be given explicit, direct controls over capacity and quality. By controlling only the price level and structure, the regulator implicitly "controls" these quality variables. Just as there is no need for capacity controls by the regulator, collusive agreements for industry "self-regulation" should be discouraged. In short, through the control of fares, the regulator has in hand an efficient means of controlling total capacity and bringing about a more efficient level and structure of price-quality options.

REFERENCES

W. J. Baumol and A. K. Klevorick, "Input Choices and Rate-of-Return Regulation: An Overview of the Discussion," *Bell J. Econ.*, Autumn 1970, *1*, 162–90.

M. Brenner, *Testimony, CAB* Docket 21866-6, Exhibit TW 6011, Washington, Aug. 1970.

R. E. Caves, *Air Transport and Its Regulators: An Industry Study,* Cambridge 1962.

A. S. De Vany, "Mathematical Theory of the Airline Firm," in D. A. Clegg et al., eds., *A Computerized Systems Planning Process for Airports and Harbors of the State of Hawaii, 1,* Santa Barbara, July 1968, Appendix B-5.

———, "The Revealed Value of Time in Air Travel," *Rev. Econ. Statist.,* Feb. 1974, *56,* 77–82.

G. W. Douglas, *Testimony, CAB* Docket 21866-9, Exhibit DOT-T-3, Washington, May 17, 1971.

———, "Price Regulation and Optimal Service Standards," *J. Transp. Econ. Policy,* May 1972, *20,* 116–27.

———, "Equilibrium in a Deregulated Air Transport Market," paper presented at Seminar on Problems of Regulation and Public Utilities, Dartmouth College, Aug. 21, 1972.

——— and J. C. Miller III, *Economic Regulation of Domestic Air Transport: Theory and Policy,* Washington, forthcoming 1974.

G. C. Eads, "Competition in the Domestic Trunk Airline Industry: Too Much or Too Little?" in A. Phillips, ed., *Competition and Regulation,* Brookings Institution, forthcoming.

———, M. Nerlove, and W. Raduchel, "A Long-Run Cost Function for the Local Service Airline Industry," *Rev. Econ. Statist.,* Aug. 1969, *51,* 258–70.

W. E. Fruhan, Jr., *The Fight for Competitive Advantage: A Study of the United States Domestic Trunk Air Carriers,* Boston 1972.

W. A. Jordan, *Airline Regulation in America: Effects and Imperfections,* Baltimore 1970.

T. E. Keeler, "Airline Regulation and Market Performance," *Bell J. Econ.,* Autumn 1972, *3,* 399–424.

J. C. Miller III, "Scheduling and Airline Efficiency," unpublished doctoral dissertation, Univ. Virginia 1969.

———, *Testimony, CAB* Docket 21866-7, Exhibit DOT-T-1, Washington, August 25, 1970.

———, "A Time-of-Day Model for Aircraft Scheduling," *Transp. Sci.,* Aug. 1972, *6,* 221–46.

H. Mohring, "Optimization in Urban Bus Transportation," *Amer. Econ. Rev.,* Sept. 1972, *62,* 591–604.

G. Renard, "Competition in Air Transportation: An Econometric Approach," unpublished M.S. thesis, M.I.T., Sept. 1970.

G. Stigler, "Price and Non-Price Competition," *J. Polit. Econ.,* Jan./Feb. 1968, *76,* 149–54.

M. R. Straszheim, *The International Airline Industry,* Washington 1969.

L. J. White, "Quality Variation when Prices are Regulated," *Bell J. Econ.,* Autumn 1972, *3,* 425–36.

J. V. Yance, "The Possibility of Loss-Producing Equilibria in Air Carrier Markets," unpublished DOT rep., Washington 1970.

———, "Nonprice Competition in Jet Aircraft Capacity," *J. Ind. Econ.,* Nov. 1972, *21,* 55–71.

Air Transport Association of America, *Major U.S. Airlines Economic Review and Financial Outlook, 1969–1973,* Washington, June 1969.

U.S. Civil Aeronautics Board, *Handbook of Airline Statistics, 1971 Edition,* Washington 1972.

———, Bureau of Economics, (1970a) *CAB* Docket 21866-6, Exhibit BE-6501, Washington, July 6, 1970.

———, (1970b) "Costing Methodology, Version 6," Washington, Aug. 1970.

———, Bureau of Operating Rights, *CAB* Docket 23852, Exhibits BOR-16, 17, Washington, March 29, 1971.

U.S. Department of Transportation, *CAB* Docket 21866-9, Exhibits DOT-D-1 through 88, Washington, Nov. 20, 1970.

Factor Proportions and Economic Rationality in Soviet International Trade 1955-1968

By Steven Rosefielde*

In 1953, Wassily Leontief made the then startling discovery that the empirical structure of American factor proportions generated in bilateral exchange with the rest of the world contradicted the theoretically anticipated result based on the Heckscher-Ohlin hypothesis. Using the logic of general equilibrium analysis in a two-factor neoclassical model, the Heckscher-Ohlin theorem predicted that America, being relatively abundantly endowed with capital, would produce export goods that intensively embodied capital and import replacement goods intensively embodying the relatively scarce factor labor. By applying the apparatus of input-output analysis, Leontief showed that paradoxically American factor proportions exhibted precisely the opposite relationship to the one anticipated on the basis of the Heckscher-Ohlin theorem. Leontief's finding initiated an avalanche of theoretical and empirical research directed towards unravelling the causal factors responsible for this unexpected and perverse result. Robert Baldwin in a recent article surveyed the evolution of the factor proportions controversy, and since all students of international trade can be expected to be reasonably familiar with the issues at stake, we shall not review them here.

It should be noted, however, and this is our point of departure, that a factor proportions study of a socialist, nonmarket, nominally centrally planned economy has never been attempted. There are several reasons why this is the case, although their cogency varies. First and foremost, until quite recently, the requisite data in input-output form has been unavailable. Secondly, even when the 1959 Soviet Input-Output table became available complete with capital and labor vectors, problems involving turnover tax distortion had to be met. Third, at the level of pure theory it could be argued that since the Heckscher-Ohlin theorem assumed perfect competition in a market environment, the study of factor proportions in a planned economic order violated a fundamental postulate of Heckscher-Ohlin analysis. Fourth, and derivatively, since central planning under Soviet socialism is generally agreed to be economically irrational when measured by the criteria of both neoclassical micro and macro theory, investigation of factor proportions generated in Soviet international trade would be a priori a waste of time as the results would surely mirror this irrationality. Fifth, and last, assuming that the computation of Soviet factor proportions was a valid exercise, the further difficulty of establishing a criterion by which the Heckscher-Ohlin rationality of Soviet factor proportions could be judged had to be met, since unlike the United States, the Soviet Union was neither unambiguously capital nor labor rich vis-à-

* Assistant professor of economics, University of North Carolina, Chapel Hill. I wish to acknowledge my gratitude to Abram Bergson, Anne Carter, John Montias, and Vladimir Treml for their valued assistance, as well as to the Ford Foundation for its financial support.

vis its trading partners. Other objections based on more recent theories of the determinants of international trade structure could also be raised, but these will be dealt with later.

Recent breakthroughs in data availability by Vladimir Treml have mitigated, if not totally obviated, the first two objections.[1] On the theoretical plane, Leontief has long since shown that the market is in no sense central to the Heckscher-Ohlin theorem, since the entire argument can readily be restated as a linear programming problem.[2] The underlying logic of factor allocation, product transformation, and commodity substitution, the core of general equilibrium theory, as a technical problem is equally amenable to perfectly competitive or perfectly planned solutions. The well-known equivalence between perfectly planned shadow prices and perfectly competitive market prices is instructive in this regard. Without contending that problems of efficient use of scarce factors are perfectly planned in the Soviet Union, following Leontief it should be obvious that factor proportions may be as meaningful for an imperfectly planned economy as for an imperfectly competitive one. Regarding the alleged irrationality of Soviet micro planning, ironically just as the Leontief Paradox pointed out the danger of abstract theorizing unsupported by empirical evidence for the American case, so our findings patently falsify the deduction that the aggregate consequence of pervasive Soviet micro inefficiency is concomitant macro-economic irrationality. Furthermore, as to the fifth objection, contrary to expectations, a measure of relative factor scarcity can be constructed so that the rationality of the structure of Soviet factor proportions is in fact susceptible to a rigorous assessment.

[1] See Treml (1964, 1967, 1968, 1972).

[2] See Leontief (1966, pp. 101–08).

One additional matter of substance must be addressed before launching into the subject at hand. Heckscher-Ohlin theory can be construed in two senses: first as a theory of efficient factor use reflecting both the micro- and macroeconomic rationality of the allocation of factors and transformation of commodities exchanged in a nation's international trade; and second as an extended theory of international factor price equalization. Heckscher-Ohlin theory as a standard of general equilibrium rationality in this paper is restricted to the former meaning alone. No judgment on the validity of the factor price equalization process is attempted.[3]

I. Basic Findings

A. *The Factor Proportions Structure of USSR-WORLD Exchange*, 1955–68[4]

Let us begin our analysis of the structural characteristics of Soviet factor proportions by considering the Soviet Leontief statistics generated in *USSR-WORLD* trade shown in Table 1.

$$\Omega = \frac{km/lm}{kv/lv} \qquad (1)$$

In this familiar embodied factor proportions ratio, the numerator is the ratio of direct-plus-indirect capital to direct-plus-indirect labor embodied in Soviet import replacement goods and the denominator the analogous value of embodied factors in Soviet exports. Factor proportions values greater than 1 imply an embodied capital intensive import bias; values less than 1 an embodied capital intensive export bias. A glance therefore at Table 1 reveals that except for the single year 1955,

[3] For a full discussion of this point, see Gottfried Haberler.

[4] Details on the scope of this study, the data utilized and computational methodology are provided in the author's book. Appendix 1 gives a highly condensed summary of these matters.

TABLE 1—THE SOVIET LEONTIEF STATISTICS GENERATED IN *USSR-WORLD* TRADE COMPARED WITH ANALOGOUSLY DEFINED AMERICAN LEONTIEF STATISTICS

Year	*USSR*	*U.S.*[a]
1947		1.2996
1951		1.0577
1955	1.0288	
1959	0.9244	
1962		1.27
1963	0.8051	
1968	0.7142	
1957–62		1.2091
1955–68	0.8681	

[a] *U.S.* data has been taken from Baldwin.

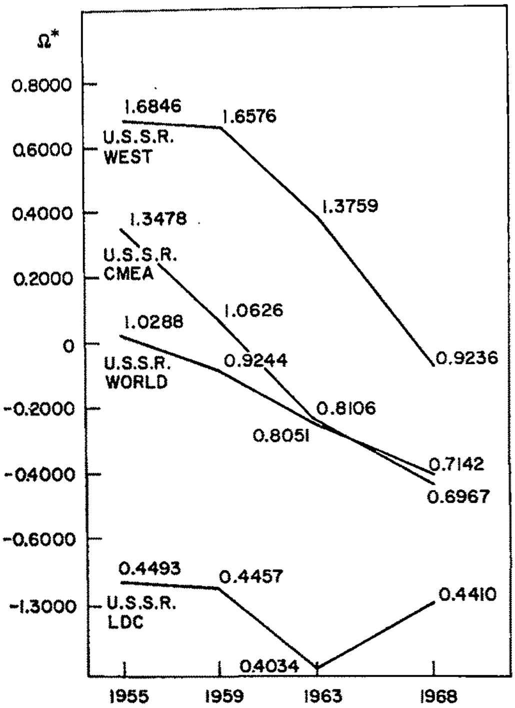


FIGURE 1. THE EMBODIED FACTOR PROPORTIONS STRUCTURE OF SOVIET TRADE WITH THE *WEST CMEA*, AND *LDC*, 1955–68

Ω^* designates the conventional Leontief statistic rescaled so that the factor proportions neutrality point is 0, rather than 1. Capital intensive import biases are then computed as $\Omega^*=\Omega-1$ while capital intensive export biases are calculated as $\Omega^*=-(1/\Omega-1)$. The reciprocal form used when $\Omega<1$ transforms the inverse numeric scaling of capital intensive export biases to the same numeric form used in scaling capital intensive import biases. As a result of this transformation, factor proportions bias is indicated by a sign convention which preserves symmetric scaling. Conventional Leontief statistics, however, are also provided directly on the figures and in the relevant tables.

the Soviet Union has consistently evidenced an embodied-capital intensive export bias vis-à-vis the world as a whole. Second, and of equal importance, by 1968 the intensity of Russia's embodied capital export bias was substantially greater than any analogous American bias. Table 1 also brings out the sharp contrast between American and Soviet factor proportions. Whereas, contrary to Heckscher-Ohlin theory, *U.S.-WORLD* Leontief statistics persistently manifest an embodied capital intensive import bias which fluctuates over time, the Soviet pattern is regular, with the factor proportions bias in the opposite direction. However, it would be invalid to infer that since the *U.S.* pattern was H-O irrational, the Soviet trend by analogy is rational, since we as yet have not determined whether vis-à-vis the world, the Soviet Union is capital or labor abundant. Before ascertaining the correct nature of relative Soviet factor availabilities, it will be useful to acquaint ourselves with the the patterns of Soviet factor proportions at a lower level of aggregation.

B. *The Structure of Soviet Factor Proportions Generated in Bilateral Exchange with the COMECON, WEST, and LDC BLOC,* 1955–68

From both the point of view of ideology and relative developmental levels, Soviet international trade can be usefully disaggregated into three familiar ideo-geographical categories: developed capitalist, developed socialist, and less developed countries. Using the criterion of ruble value of foreign trade turnover, England, West Germany, France, and Finland were selected to represent the first or *WEST* classification; East Germany, Czechoslovakia, Poland, Hungary, Bulgaria, Rumania, the second or Council for Mutual

TABLE 2—SOVIET LEONTIEF STATISTICS GENERATED IN TRADE BETWEEN THE *USSR*, *CMEA*, *WEST*, AND *LDC*[a,b]

USSR Trade Pattern With:	Soviet Leontief Statistics				
	1955	1959	1963	1968	1955–68
CMEA	1.3478	1.0626	0.8106	0.6967	0.9794
WEST	1.6846	1.6576	1.3759	0.9236	1.4400
LDC	0.4493	0.4457	0.4034	0.4410	0.4349

[a] The general formula for Soviet Leontief statistics computed for *USSR*-Regional or *USSR*-National trade is

$$\Omega_{ij} = \frac{k_{59}m_{ij}/l_{59}m_{ij}}{k_{59}v_{ij}/l_{59}v_{ij}} \qquad \begin{array}{l} j = j(1, \ldots, 17) \\ i = i(1, \ldots, 4) \end{array}$$

where k = direct-plus-indirect capital requirements per ruble of sectoral output
l = direct-plus-indirect labor requirements per 10,000 man years per ruble of sectoral output
m = import competing replacement goods
v = exports
i = the year index
j = the country or regional index

[b] Only the European members of the *CMEA* excluding Albania are included in this study; *WEST* is defined as the United Kingdom, West Germany, France, and Finland which on average 1955–68 account for 53 percent of Soviet trade with developed capitalist countries; *LDC* is defined as China, India, and the *UAR*.

Economic Assistance (*COMECON* or *CMEA*) category; and China, India, and *UAR*, the last or *LDC* group.[5] Note that from the Soviet standpoint, our *LDC* unit consists of less developed countries of both capitalist and socialist varieties. Disaggregated sectoral analysis shows, however, that the factor proportions structure generated in trade with the *LDC*s is basically the same regardless of national ideology.

All the relevant data required for analyzing the bloc structure of Soviet Leontief statistics is contained in Table 2 and represented graphically in Figure 1. The regularity of the bloc factor proportions pattern is so stark that our results are almost self-explanatory. The relative magnitude of the *USSR-WORLD* Leontief statistics decomposes into three discrete sets. On average, trade between the *USSR* and the *WEST* is intensively capital import biased, between the *USSR* and the *LDC* intensively capital export biased, and between the *USSR* and the *CMEA* factor neutral. Since by most standards of international comparison the *WEST* is more developed than the *LDC*, the direction of embodied factor intensive bias is positively correlated with the level of development of the trading bloc. Thus when the Soviets trade with relatively developed countries, they import capital intensive commodities and export labor intensive goods. When they trade with nations at a similar developmental level, imports and exports embody factors in approximately equal proportion, and when they trade with underdeveloped regions,

[5] The selection of the thirteen countries was based on two stringent criteria: ruble volume of trade turnover and its continuity for all the years sampled. Since on average these thirteen countries account for approximately 80 percent of Soviet trade, and the time cost of the basic data preparation is high, this study was not expanded to include additional, marginal observations. Nevertheless, as a check *USSR*-Cuba Leontief statistics were computed for 1963 (0.5819) and 1968 (0.6589). If the 1968 Cuban Leontief statistic value is plotted in Figure 4, given Cuban per capita *GNP* of $900 in 1967 prices, the reader can readily ascertain the basic conformity of this latest observation with the original thirteen.

they export capital intensive and import labor intensive products.

The average pattern, however, is only half the story. The time trends are also significant. With the exception of the *USSR-LDC* case, the period 1955–68 is characterized by a dramatic change in the magnitude and direction of the *USSR-WEST* and *USSR-CMEA* factor proportions bias. In just thirteen years we encounter a factor bias reversal transforming the Soviet Union from a labor intensive to a capital intensive exporter of embodied factors to the *WEST*. Similarly, *USSR-CMEA* trade exhibits a significant factor bias reversal between 1955 and 1963. These trends serve as an intriguing index of the extent to which the structure of the Soviet economy has been transformed in recent years.

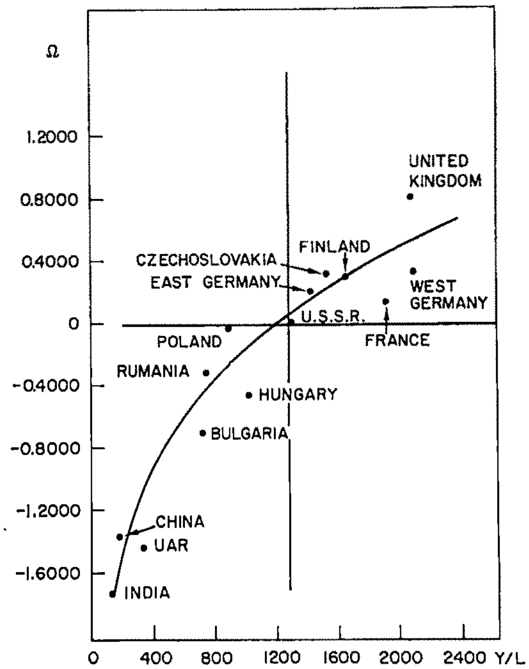


FIGURE 2. SOVIET LEONTIEF STATISTICS PLOTTED AS A FUNCTION OF AVERAGE PER CAPITA *GNP*, 1955–68

II. A Method For Evaluating the Compatibility of Soviet Leontief Statistics with Heckscher-Ohlin Theory

To test the hypothesis that Soviet Leontief statistics are generated according to the principles embodied in the Heckscher-Ohlin theorem, we hypothesize that there exists a correlation between the factor availability ratios of the Soviet Union's trading partners and the value of the Soviet Leontief statistic.

$$\Omega = \phi\left(\frac{K}{L}\right) \tag{2}$$

where Ω = the Soviet Leontief statistics, K/L = the capital-labor ratio. If the trading partner is capital rich vis-à-vis the *USSR*, we would anticipate high values for the Soviet Leontief statistic reflecting the importation of capital intensive and the exportation of labor intensive commodities. Where the trading partner is labor rich, the reverse result would be expected. Since direct measures of the requisite capital-labor ratios are unavailable, per capita *GNP* was utilized as a proxy in order to quantify Soviet factor endowment ratios relative to its trading partners. Although many functional specifications linking per capita national income and the capital-labor ratio are plausible, the following form derived from the Cobb-Douglas aggregate production function yielded very satisfactory results:

$$\Omega = \tau + \beta \log \frac{Y}{L} \tag{3}$$

where Y/L signifies per capita *GNP*.[6]

Let us consider the relationship between average Soviet Leontief statistics and average per capita *GNP*. Figure 2 illustrates this relationship for trade between the *USSR* and the thirteen coun-

[6] For a more detailed description of these matters, see the author.

TABLE 3—PER CAPITA GROSS NATIONAL PRODUCT IN 1967 DOLLARS WITH CONSTANT 1955 EXCHANGE RATE ADJUSTED TO A GEOMETRIC MEAN PURCHASING POWER BASIS

Country	Average	1955	1957	1959	1963	1965	1968
U.S.	3513	3145		3195	3487		4226
WEST	1937	1562		1748	2044		2395
United Kingdom	2094	1821		1970	2164		2424
West Germany	2093	1588		1897	2234		2652
France	1906	1493		1672	2010		2450
Finland	1655	1346		1453	1769		2052
COMECON[a]	1052	798		952	1096		1363
Czechoslovakia	1536	1198		1456	1576		1912
East Germany	1412	1041		1358	1500		1748
USSR	1278	975		1182	1309		1649
Hungary	1011	787		888	1062		1308
Poland	883	715		792	899		1124
Rumania	741	564		612	764		1024
Bulgaria	731	485		606	773		1061
LDC[b]	244	198		175	257		262
UAR	319	252		299	348		357
China	190		190			190	
India	157	144		150	166		166

[a] *COMECON* (Council for Mutual Economic Assistance) excludes the *USSR*.
[b] *LDC* excludes China. If China is included, the average per capita *GNP* is 224.

tries selected for this study.[7] Soviet Leontief statistics are arrayed along the ordinate; per capita *GNP* along the abscissa. The diagram is divided into four quadrants by making the hypothetical observation of the *USSR-USSR* trade the origin of a set of coordinate axes. By inspection we find that countries with per capita *GNP* less than the Soviets' all fall in the third quadrant, while those possessing per capita *GNP* greater than the *USSR* fall entirely in the first quadrant. This suggests that there exists a tendency for Soviet Leontief statistics to be positively associated with our per capita *GNP* proxy for the capital-labor availabilities ratio. Put somewhat differently, in accordance with Heckscher-Ohlin theory, the Soviets export relatively labor intensive goods to capital rich countries and relatively capital intensive ones to labor abundant nations. The clustering of points in quadrants I and III in Figure 2 demonstrate that Soviet trade with the member states of the *CMEA*, *WEST*, and *LDC* conforms with the principles of Heckscher-Ohlin theory insofar as per capita income validly reflects factor endowments. It is equally evident that the semilog specification of equation (2) describes the observed scatter exceedingly well. A semilog regression run on all fourteen country observations, including the hypothetical *USSR-USSR* point, yields an R of 0.9615 with a β coefficient of 1.98004 significant at the 0.995 level. Since our observations come from a cross-section study, rather than a time-series, this semilogarithmic fit explaining 92.5 percent of the variance in the dependent variable must be viewed as a very impressive indication of the basic conformity of Soviet factor proportions with the laws of Heckscher-Ohlin theory. This conclusion is further buttressed by noting in Figure 2 that the semilogarithmic regression line passes very close to, if not through, the *USSR-USSR* point, showing that there does not exist any unexplained

[7] Per capita *GNP* values were computed on a purchasing power parity, geometric mean basis in 1967 dollars. For details, see the author. For the values themselves, see Table 3.

cumulative capital intensive import or export bias. These are exceedingly important findings not only because they represent the fullest possible test of the Heckscher-Ohlin consistency of Soviet factor proportions in trade with the countries studied, but the semilogarithmic fit suggests that our per capita *GNP* variable is consistent with a Cobb-Douglas production function specification, attesting significantly to the conclusion that the Heckscher-Ohlin theorem consistency of our Soviet Leontief statistics is not a mere fluke, but reflects valid underlying economic forces.

The strongly Heckscher-Ohlin rational results obtained thus far have been based on average values for Soviet Leontief statistics and per capita *GNP*, 1955–68. In so far as the averaging process has tended to eliminate random fluctuations in these variables in any specific year, the outcome has been to the good. But in our analysis of Table 2 we detected a secular trend towards an increased capital intensive export bias in Soviet trade with the *CMEA* and the *WEST*. Since the averaging process conceals the impact of this secular trend, it is now necessary to discover what effect this secular trend has had first on the validity of the Heckscher-Ohlin theorem as an explanation of Soviet Leontief statistic values, and second on the functional form of the relationship between the observed Soviet Leontief statistics and per capita *GNP*. We will endeavor to assess the consequence of the secular trend by examining the correlation between Soviet Leontief statistics and per capita *GNP* in 1955 and 1968. The initial and terminal years of our study have been selected for special consideration because they demonstrate the maximum impact of the secular trend on the Soviet Leontief statistics. Figure 3 illustrates the point scatter of Soviet Leontief statistics and per capita *GNP* for the year 1955. We note first as might be expected that the observations are obviously more widely dispersed than in Figure 2, due in part to random forces at work in any single year. Nevertheless a semilog specification appears to fit the scatter reasonably well, although the slope is somewhat steeper than the gradient in Figure 2. Secondly, we must observe that two points, one representing *USSR*-Rumanian and the other *USSR*-Polish trade, fall in quadrant II. This indicates that the Soviet Leontief statistics associated with these points are, contrary to Heckscher-Ohlin theory, negatively correlated with our capital-labor availabilities proxy. On the whole, however, the positive correlation required prevails with a slight tendency evident for the Soviet Leontief statistics to exhibit labor intensive export biases somewhat greater than called for on the basis of Heckscher-Ohlin theory. Thus, Figure 3 shows that the averaging process for the initial year 1955 conceals a modest

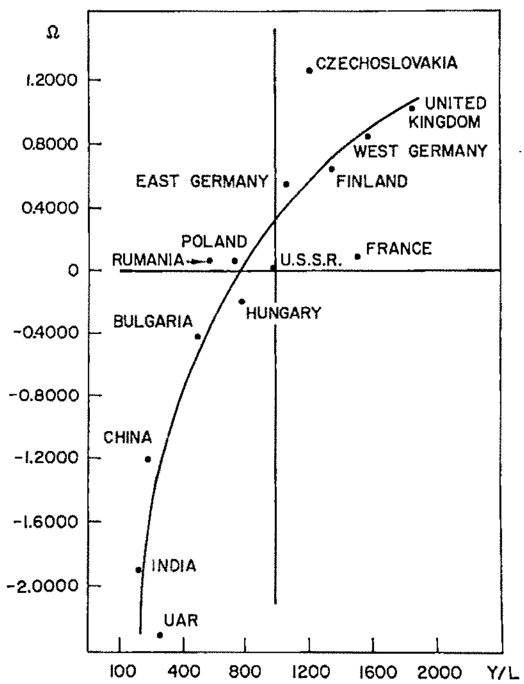


FIGURE 3. SOVIET LEONTIEF STATISTICS PLOTTED AS A FUNCTION OF 1955 PER CAPITA *GNP*

labor intensive export bias, but the functional form remains semilogarithmic and broadly compatible with Figure 2.

When we turn our attention to the relationship between Soviet Leontief statistics and per capita *GNP* in 1968, however, the situation is considerably altered. Figure 4 suggests that the functional form of the relationship between Soviet Leontief statistics and per capita *GNP* is still semilogarithmic, but the regression line passes to the southeast of the *USSR-USSR* trade point indicating a secular capital intensive export bias, expressed as a parametric shift factor in *USSR-CMEA* and *USSR-WEST* commodity trade. As a consequence of this secular trend, five observations all for the relatively developed nations of the group fall in quadrant IV showing a negative correlation between these Soviet Leontief statistics and per capita *GNP*. In 1955 the negative correlation indicated a violation of the Heckscher-Ohlin theorem with a bias to excessively labor intensive exports; in 1968, however, the negative correlation refers to a bias towards excessively capital intensive exports. Both the number of negatively correlated observations and the magnitude of their deviation from their correct location in quadrant I is greater in 1968 than in 1955. If the secular trend displayed over the period 1955–68 continues, we would anticipate that sometime in the near future the Soviet Union will demonstrate a capital intensive export bias in its trade with all countries. Some of these capital intensive export biased points in quadrant III will be justified according to Heckscher-Ohlin theorem, but the points in quadrant IV will violate the theorem even though a positively sloped semilogarithmic curve adequately describes the functional relationship between the Soviet Leontief statistics and the capital-labor availabilities proxy. Thus, Figures 2 and 4 taken together reveal that the averaging of the values of both variables 1955–68 conceals a tendency for Soviet Leontief statistics to be progressively less compatible with the principles of Heckscher-Ohlin theory, due to a secular trend towards an overall capital intensive bias of Soviet exports with all trading partners, regardless of their level of economic development. Nevertheless, despite the excessive capital intensity of Soviet exports to the *WEST* and certain *CMEA* members, a positive, though weakening, semilogarithmic correlation between Soviet Leontief statistics and per capita *GNP* still persists indicating that the factor endowments of the *USSR*'s trading partners is not insignificant.

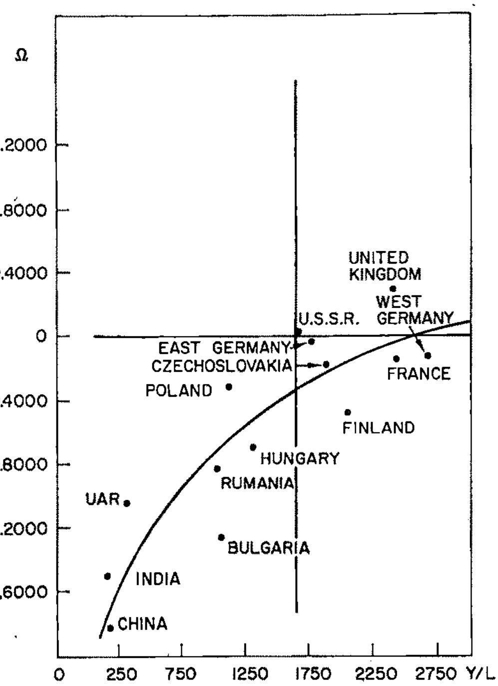


FIGURE 4. SOVIET LEONTIEF STATISTICS PLOTTED AS A FUNCTION OF 1968 PER CAPITA *GNP*

III. Some Tentative Explanations of the Heckscher-Ohlin Rationality of Soviet Factor Proportions 1955–68

Regardless of the empirical nuances, one overriding theoretical conclusion stands out starkly from the statistical detail of

our study: embodied Soviet factor flows conform with extraordinary fidelity to the strictures of Heckscher-Ohlin theory. Whatever special ideological forces may be at work in determining the composition of Soviet trade with the Socialist bloc, they are seemingly subordinate to the dictates of pure theory. Just as in the case of the Leontief Paradox, a result which so flagrantly flouts conventional wisdom begs for an explanation. Aggregation experiments have shown that our findings are not seriously impaired if we reduce the Soviet Input-Output matrix from sixty-six to four sectors: agriculture, light industry, heavy industry, and natural resources. A time-series study of Soviet commodity structure utilizing these four categories demonstrates conclusively that Soviet factor proportions trends reflect the burgeoning importance of consumerism in Russia. Unable to satisfy domestic consumer demand internally, Soviet policy makers have systematically expanded light industrial imports which have been paid for more and more by natural resource and heavy industrial exports. Since light industrial goods are relatively labor intensive while natural resource and heavy industrial goods are strongly capital intensive, the politically motivated shift in the structure of Soviet commodity trade is undoubtedly correlated with the progressively increasing capital intensity of Soviet exports and the concomitant increase in the labor intensity of Soviet imports noted in Section II.

Clearly, there exists a political dimension to the rationality of our economic results. But this cannot be the whole story. What is missing is the establishment of a plausible mechanism which explains how Soviet planners systematically arrange the commodity composition of traded goods so that Soviet Leontief statistics are semi-logarithmically correlated with the per capita *GNP* of its trading partners. Several possibilities can be entertained ranging from perfect planning to the semi-divine intervention of Stalin's not too invisible hand. Attention here is focused on what the author believes to be the least implausible of a wide variety of implausible explanations of the Heckscher-Ohlin rationality of Soviet foreign trade. Specifically, after sorting through the available alternatives, it appears that the Soviet results are the outcome of comparative advantage, labor value, accounting price calculations.

Soviet price formation is directly linked to factor costs. If factors have a definite structure of over- or undervaluation, the possibility arises that commodity prices will be regularly biased to the extent that they embody more of one or the other factor. Commodities intensively embodying the overvalued factor will appear relatively dear in comparison with commodities intensively embodying the undervalued factor. If the principles of price formation function in this manner, we might anticipate seeing evidence of it in the embodied factor content of Soviet trade. An undervalued factor intensive export bias should occur measured against the Soviet Union's actual factor endowment ratio. To ascertain whether this is indeed the case, let us briefly examine the principles of Soviet price information.

Soviet commodity prices are formed on the basis of the labor theory of value. Before the 1967 reforms, labor value accounting prices were broadly determined by average direct-plus-indirect labor costs prevailing in a particular branch of production. Indirect labor alludes to the embodied labor content of material inputs and capital employed in each branch of production. A depreciation charge was also set.[8] As is widely understood, accounting prices of this sort diverge from their

[8] See Morris Bornstein.

scarcity counterparts due to the omission of interest and rent charges. Capital goods as a consequence will in general be priced below their opportunity cost. This undervaluation of capital has a cumulative dimension as well since the value of the capital good employed by a given branch will be understated by analogous omissions of rent and interest charged on the capital used to produce the second generation of capital goods in question. With time, therefore, goods produced in capital intensive industries should appear to be domestically cheap measured by prevailing accounting labor value prices. But this process cannot possibly determine the pattern of Soviet foreign trade in and of itself. From the perspective of international exchange, relative domestic dearness or cheapness is important only in comparison with the ruling international terms of trade. Purchasing power parity relatives derived by Treml from Efimov's computations shed light on this important issue. Partitioning Soviet output into two categories, producers' goods and consumers' goods, which accord roughly with the capital intensive, labor intensive distinction relevant in Heckscher-Ohlin analysis, purchasing power relatives indicate that in terms of Soviet accounting prices, producers' goods are relatively less expensive than consumers' goods in comparison with international prices. Consequently, according to the principle of comparative advantage, Soviet prices signal the system directors to export capital intensive producers' goods in exchange for labor intensive consumers' goods. If this conjecture is a correct one, and it is remembered that Soviet accounting procedures tend to understate the cost of capital intensive goods, a testable null hypothesis suggests itself: the embodied factor proportions pattern of Soviet foreign trade does not exhibit capital intensive export biases in bilateral exchange between the Soviet Union and countries relatively more abundantly endowed with capital.

Consider Figures 2 and 4 once again. Since theory at this point is an imperfect guide, in testing the null hypothesis we seek only to discover whether observed bilateral Soviet Leontief statistics are distributed in the capital intensive export bias quadrant, that is, quadrant IV. Over the entire period 1955–68 on average, Figure 2 reveals that the null hypothesis is validated. No observations appear in quadrant IV, indicating that the Soviets do not import relatively labor intensive goods from countries that are relatively capital rich. If labor value accounting prices prompt the system directors to import an excessively labor intensive bundle of goods from the *WEST*, there is no discernible evidence of it. The *USSR-WEST* Leontief statistics are clearly compatible with Walrasian price expectations in the aggregate, even if prices fail to conform to the consumer's utility standard in a more detailed way. However, the average values misrepresent the trend. Figure 1 demonstrates that between 1955 and 1968 *USSR-WEST* trade becomes steadily less capital intensive import biased until, in 1968, the capital bias is transformed into a substantial labor intensive import bias. The *USSR-CMEA* exchange follows the same pattern. By 1968, Figure 4 reveals Soviet trade is labor intensive import biased with five of its thirteen trade partners; five Soviet Leontief statistics fall in quadrant IV. The null hypothesis is no longer sustainable. Although it cannot be demonstrated within the confines of this article, neither natural resources nor labor skills are primarily responsible for this outcome.[9] The new pattern of *USSR-WEST* trade in the aggregate now conforms with labor value comparative advantage, rather than Walrasian principles. Furthermore, compari-

[9] See the author.

son of Figures 2 and 4 reveals a decisive increase in the labor intensive import biases generated in *USSR-CMEA* trade from average, 1955–68 Soviet Leontief statistic values, and those for 1968 alone. Pending a more cogent interpretation, the least implausible explanation of the empirical data at hand appears to suggest that the Soviets have relied on labor value prices as a guide to trade decision making in a much more profound way than is commonly understood. Although the evidence is really insufficient to draw anything but a tentative conclusion, the persistence of the Leontief statistic trends uncovered in this study beyond 1968 will serve as a fundamental test of our labor value comparative advantage conjecture. Validation will not only illuminate the deeper significance of our Heckscher-Ohlin consistent results, but should lead to a reassessment of the whole question of rational foreign trade planning under Soviet socialism.

One final word is in order. In the introduction it was pointed out that the microeconomic irrationalities of the Soviet economic system would lead one to anticipate a macro-economically irrational, non-Heckscher-Ohlin structure of embodied Soviet factor proportions. As we have shown, if anything the reverse is true. This suggests that in evaluating the economic merit of alternative economic systems, micro-economic reasoning may well serve as a faulty guide to assessing macro-economic consequences of micro-economic decision making.

It would seem that further research on this complex theme is required if the field of comparative economic systems is to come to grips with the more profound aspects of the general systems problem.[10]

[10] For an elaborate exposition of this subject, see Janos Kornai.

Appendix

A. *Data*

Data of two sorts underlie the whole of the foregoing analysis. First, we utilize standard Soviet foreign trade statistics. Second and more importantly we rely on the 66 x 66 square 1959 Soviet Input-Output table as reconstructed by Treml as the statistical base for interindustrial flow, final demand, import, export, labor, skill, and capital stock data. In general this table conforms closely with the Leontief-type, open, static, four-quadrant Input-Output table familiar in the West. Without elaborating its detailed content, however, several salient aspects peculiar to the Soviet table which could potentially impair its comparability with western tables, and by extension qualify the meaning of our factor proportions analysis require explicit acknowledgement. First, and most significant, instead of employing factor cost prices to value physical flows, the Soviets used purchasers' prices. These prices fall short of the theoretical norm not only by the inclusion of turnover taxes and the omission of charges for interest and rent, but in addition by the inclusion of trade and transportation costs paid by purchasing firms which double counts the same charges already recorded in the appropriate transportation and distribution vectors. Second, following the Soviet definitions of national income, some unproductive activities such as health and financial services are not separately identified in the table, while others, notably passenger transportation and communications serving the public, are treated as claimants against final demand. Third, and to the good, secondary products in the industrial sectors have been carefully reallocated to the industries in which they are primary so that each activity is relatively pure. Finally, the Soviet capital-output vector employs an observed rather than a capacity measure of output which means that insofar as observed sectoral production falls short of capacity output, the relevant Soviet capital output ratios overstate incremental direct-plus-indirect capital requirements. Additional details on both the con-

struction and reconstruction of the 1959 Soviet Input-Output table can be found in Treml's numerous publications. For our purposes, suffice it to say, that given the quality of the basic Soviet data inputs, including a 20 percent random stratified industrial survey and the gigantic 1959 economywide inventory and revaluation of capital, the 1959 Soviet table is sufficiently reliable to justify our study.

B. *Turnover Tax*

Although double counting poses a genuine problem, the crucial defect of purchasers' prices is the effect of turnover taxes which distort intersectoral physical flows by differentially valuing equivalent physical volumes in alternative use. In order to overcome this troublesome problem, we have attempted, albeit crudely, to remove turnover tax from the 1959 Soviet Input-Output table. Using purchasing power parity ratios derived by Treml from Efimov's computations, an estimate of the ratio of final to intermediate purchasers' prices was obtained for three aggregate sectors, producers' goods, consumers' goods, and agricultural commodities which enabled the consistent valuation of the entire Input-Output table in intermediate purchasers' prices. This deflation procedure potentially eliminates 80 percent of the distortion introduced by the turnover tax and greatly enhances the reliability of our findings.

REFERENCES

R. Baldwin, "Determinants of the Commodity Structure of U.S. Trade," *Amer. Econ. Rev.*, Mar. 1971, *61*, 126–46.

M. Bornstein, "The Soviet Price System," *Amer. Econ. Rev.*, Mar. 1962, *52*, 64–103.

G. Haberler, *A Survey of International Trade Theory*, Princeton 1965.

J. Kornai, *Anti-Equilibrium*, Amsterdam 1972.

W. Leontief, "Domestic Production and Foreign Trade: The American Capital Position Re-examined," *Proc. Amer. Philosophical Soc.*, Sept. 1953, 332–49.

———, "Factor Proportions and the Structure of American Trade: Further Theoretical and Empirical Analysis," in *Input Output Economics*, New York 1966.

S. Rosefielde, *Soviet International Trade in Heckscher-Ohlin Perspective: An Input Output Study*, Lexington 1973.

V. Treml, "The 1959 Soviet Intersectoral Flow Table, Volume I," tech. pap. RAC-TP-137, Research Analysis Corp., Nov. 1964.

———, "New Soviet Capital Data," *Soviet Stud.*, Jan. 1967, *18*, 29–295.

———, "New Soviet Inter-Industry Data" in *Recent Soviet Economic Performance: Selected Aspects*, tech. pap. RAC-P-38, Research Analysis Corp., Aug. 1968.

——— et al., *The Structure of the Soviet Economy: Analysis and Reconstruction of the 1966 Input-Output Table*, New York 1972.

The Interaction Between Local Government and Urban Residential Location

By RICHARD E. SCHULER*

What is an appropriate way to allocate public services within an urban area? Attempting an answer requires a definition of urban public services. Many are intended to offset externalities that are due to crowded living patterns in urban areas. As an example, municipal water supply, sewage disposal, and garbage collection are all attempts to ameliorate unpleasant realities of human existence. In rural areas, the solution is left to individual households, partly because this approach is less costly with widely separated homesites, but also because an unsuccessful solution imposes a much smaller burden on distant neighbors.

Similarly, if dense habitation breeds antisocial behavior, then police and a system of jurisprudence can be thought of as attempts to offset this externality. That portion of schooling designed to socialize the human species would also fall in this category and therefore would be of greater relative importance in urban areas. Fire can have a devastating effect on a rural household, but its probability of spreading into a conflagration that will harm many individuals is minimal outside of urban areas. Finally, noise and residentially generated air pollution are both emphasized by residential crowding.

Despite these adversities we observe households moving to densely inhabited urban areas. They gain higher incomes which result from both internal and external scale economies of production. Since accessability to a wide variety of economic agents is a primary force underlying those economies in urban areas, households accept dense habitation patterns in order to improve their accessability and incomes. But in doing so, they also accept the offsetting negative externalities created by density.[1]

This analysis will deal only with density-offsetting urban public services. The urban transportation system will be assumed as given, although changes in accessibility certainly will alter urban shape and growth.[2] Finally, welfare payments and that portion of education designed to increase skill levels will not be considered here, partially because income transfers and human capital improvement are not quite so uniquely an urban concern.

* Assistant professor of economics and environmental engineering, Cornell University. I want to thank Martin Beckmann for laying the foundation for this analysis and for his incisive comments. Encouragement and many helpful suggestions were received from Benjamin Chinitz and participants in his urban seminar at Brown University and from members of George Borts' thesis workshop also at Brown. I am also grateful to Robert Rohr, Harl Ryder, Henry Wan, Jr., and Erwin Blackstone for their useful comments, An early version of this paper was presented at the annual meeting of the Regional Science Association, November 1971, at Ann Arbor, Michigan.

[1] One might argue that human interaction and therefore dense habitation is a basic human desire. This may be so up to a point, but if the tendency is unbounded, what is to prevent households from achieving infinite densities by avoiding commuting costs and living in the Central Business District (*CBD*)?

[2] Robert Solow has already studied the optimum allocation of land for roads within an urban, spatial equilibrium framework.

Now, reconsider the original question of how to allocate urban public services, and focus on those that offset externalities which arise from population density. An egalitarian approach under the assumption of neutral financing would be to distribute these services equally. But does this mean, as an example, one should have the same number of policemen per capita patrolling in every neighborhood, or should we attempt to equalize the crime rate? The second objective implies a much higher per capita concentration of policemen in densely populated areas. But this egalitarian approach runs counter to the price theorist's prescription for dealing with externalities.[3] He would argue that if dense habitation generates negative externalities, then dense habitation should be discouraged. An efficient solution would be to tax dense habitation or to place more policemen per capita in the suburbs. Either policy would encourage sparse settlement patterns. The problem with this usual economic analysis when applied within an urban context is that it neglects the effect of policy upon commuting. As an example, in a central business district (*CBD*) focused city, increased surburbanization implies higher commuting costs which may be viewed as a dead weight loss to society.

I. Relationships Between Local Government and Residential Location

If one is interested in the long-run consequences of any policy to allocate these urban public services, the possibility of household relocation must be considered. Unfortunately, many interrelationships exist between government activity and decisions by households on where to locate within an urban area. Charles Tiebout hypothesized that households consider urban differences in the public service-revenue source mix in choosing a location. Wallace Oates has provided empirical verification of this hypothesis. But there are also feedback effects. Residential location patterns have an impact on both the distribution and aggregate level of land rents in any urban area, and capitalized land rents represent an important portion of the urban property tax base. By influencing the tax base, local government may unknowingly precipitate its own fiscal crisis or bonanza by the way it distributes public services, unless it foresees this relocational effect. It may also invalidate its original prescription for allocating public services. Although the effect will not be considered here, population density patterns may also influence the production function for many public services.[4] Since approximately 20 percent of the residents of a metropolitan area move each year and three-quarters of these relocate in the same area,[5] local government's capacity to rapidly alter the shape of an urban area may be very real.

By comparison, traditional public finance theory has focused on determining an optimal bundle of public goods and an appropriate scheme for raising the required revenue with the size and distribution of the population to be served as given and invariant parameters. As an example, Paul Samuelson has developed the marginal conditions associated with the optimum bundle of his definition of "pure public goods," but here the level of "public goods" and their cost is assumed to be independent of the population served. James Buchanan's Club Theory is more applicable to local government since it allows some individuals to be excluded from the benefits of government expenditure, but neither theory explicitly considers the impact of these public service de-

[3] See William Baumol for a recent example of this type analysis.

[4] See James Ohls and Terence Wales for a recent study suggesting this is not an important effect.

[5] See John Lansing and Eva Mueller.

cisions on residential location. Similarly, tax incidence theory, whose state-of-the-art was recently summarized by Peter Mieszkowski, takes the bundle of public services to be provided as given. It does permit inferences about relocation to be made on the basis of income and price effects, but the further feedback ramifications on the optimum level of public services are not considered.

Most theoretical models of residential location in urban areas[6] are also relatively unsophisticated. They are static and assume a *CBD* focused shape in which all economic activity transpires downtown. Under the further simplifying assumption of identical household utility functions and a specific income distribution, the properties of a residential spatial equilibrium have been investigated by Muth, Martin Beckmann, Aldo Montesano, Solow, and A. Mitchell Polinsky and Steven Shavell.[7] The advantage of these spatial equilibrium models is that a public sector can be readily included, and they provide a framework to analyze the long-run effects of the many ramifications of government policy.

Thus a simplistic shape will be assumed for urban areas in order to gain qualitative insights into a complex set of interactions. So long as external economies of agglomeration encourage some clustering of economic activity (shopping centers, research and industrial parks, etc., as current examples) the ensuing analysis should have some applicability. First, the locational choice model for households will be summarized. The properties of a spatial equilibrium will be reviewed for a city where all households have the same incomes and tastes. Within this framework, the influence of public policy will be analyzed with respect to both the allocation of services and the selection of revenue sources. Finally, the analysis will be repeated for a community where incomes vary with some known distribution, and the appropriate public policies will be compared.

II. Model of Household Equilibrium

Consider households who seek to maximize a two-characteristic utility function of Cobb-Douglas form.

$$U = U_1 x^{1-c} Q^c \tag{1}$$

where Q = Level of social environment in one's neighborhood
x = Composite of all other consumption goods
U_1, c = Constants

Here, the level of social environment is a utility-enhancing characteristic in the Lancaster sense. It declines as the population density in one's neighborhood increases, but it can be augmented by the provision of public services as shown in (2).[8]

$$Q = Q_o\left(\frac{w}{\phi}\right) \tag{2}$$

where w = Level of public services per household
ϕ = Population density
Q_0 = Constant

Within this framework, w would be similar to policemen per capita in a specific neighborhood, as an example; whereas Q^{-1} would be analogous to the resultant crime rate. This formulation appears realistic since the only case where a positive level of social environment, and therefore of utility, can be maintained without any public services is in a rural area where

[6] See William Alonso, Edwin Mills, Lowdon Wingo, and Richard Muth for examples.

[7] A spatial equilibrium may also be easily demonstrated for the community with equal incomes where tastes vary with some known distribution (see the author.)

[8] Realistically, w and ϕ in equation (2) should be raised to different exponents; however, it can be shown that the following analysis does not vary qualitatively as a result of the simplifying assumption.

population density also approaches zero.

Let local government be free to allocate public services in accordance with population density.

$$w = AW_{tot}\phi^{g} \tag{3}$$

where W_{tot} = total quantity of public goods provided in the community, and A, g = Constants. Here the important policy variables are W_{tot} and g. If g is negative, the distribution of public services will be skewed in favor of sparsely settled areas. If g is positive, dense areas will be favored. Where $g=0$, the level of public services per household is equal everywhere in the city. The value of the constant A will be determined by the spatial equilibrium analysis since it must insure that the sum of all per-household public services equals the total.

Now, combine (2) and (3) in (1).

$$U = U_1[Q_oAW_{tot}]^c x^{1-c}\phi^{c(g-1)} \tag{4}$$

Under this formulation the individual household recognizes how population density affects its well-being in two ways: directly by influencing the level of social environment, or indirectly through local government's policy of distributing its services. Households are not aware of how their choice of a neighborhood on the basis of its density might affect W_{tot}, the total quantity of public services provided throughout the city.

If an argument similar to Homer Hoyt's (that households with similar tastes and incomes will choose to live in the same neighborhood) is made, then all households with the same incomes will demand identical quantities of the same goods, and they will want to live at the same distance from the *CBD*. Since tastes do not vary, land occupied per household will be identically equal to the inverse of household density. The utility function and budget constraint may be reformulated in the following familiar Lagrangian form:

$$H = U_o x^a s^b + \lambda[y - xP - sR(r) - T_o r] \tag{5}$$

where

s = Space occupied per household = $1/\phi$
y = Net household income
P = Price of the composite good
R = Land rent per unit area
r = Distance from the *CBD*
T_o = Out-of-pocket commuting cost per unit distance
$U_o = U_1[Q_oAW_{tot}]^c$
$a = 1-c$
$b = c(1-g)$
$n = a+b = 1-cg$

Distance from the *CBD* could enter the utility function negatively since commuting reduces leisure time and/or because it is a nuisance. Conversely, distance could enhance utility, as argued by Polinsky, because the pollution and noise generated by the economic activity in the *CBD* is dissipated only gradually over distance. These two views tend to offset each other, and for simplicity distance has not been included in the utility function.

All households are assumed to face the same prices, including out-of-pocket commuting cost per mile. This implies that each household makes the same number of trips to the *CBD* each week. If publicly provided, commuting is assumed to be completely supported by user charges. Household income should affect per mile out-of-pocket commuting costs only if a viable income-earning opportunity is foregone. This may not be the case for many Americans who work a "normal" forty-hour week.[9]

[9] A primary reason for focusing on a commuting cost form which is only a function of distance is that for homogeneous utility functions, the sign of the total derivative of distance with respect to income hinges uniquely on it. As an example, if commuting costs are of the form $T_1yr^a+T_0r^b$, then it can be shown that $dr/dy>0$ so long as $T_0>0$. If $T_0=0$, dr/dy is indeterminate, and an equilibrium rent gradient can be constructed for a variety of assumed locational patterns by income.

Given the vector of prices and its income, each household in equilibrium will seek to maximize equation (5) by choosing x, s, r and λ. The first-order conditions may be solved explicitly, and they yield the following well-known demand equations.[10]

$$x = \left(\frac{a}{n}\right)\left(\frac{y - T_o r}{P}\right) \tag{6}$$

$$s = \left(\frac{b}{n}\right)\left(\frac{y - T_o r}{R}\right) \tag{7}$$

$$s = \frac{-T_o}{\frac{dR}{dr}} \tag{8}$$

It can be shown that the wealthy will always live further from the *CBD* under this formulation.[11]

III. The City's Households Have Equal Income

A. *Spatial Equilibrium*

All of the individual's equilibrium conditions must be met in a spatial equilibrium. While each household is assumed to know the rent gradient confronting it, the problem now is to construct that equilibrium rent gradient for the city, given its population, available land, and aggregate level and distribution of household income. Since the supply of all other goods, x, is assumed to be perfectly elastic, only two additional conditions are imposed: the supply and demand for land must be equal at every point throughout the city, and all households must be assigned a location within its boundaries.

The equilibrium rent gradient is easily determined for the equal income city by combining (7) and (8).

$$\frac{\frac{dR}{dr}}{R} = \frac{-T_o\left(\frac{n}{b}\right)}{(y - T_o r)} \tag{9}$$

[10] See Solow for these exact results.

[11] See Muth and the author.

Here income is a constant and the solution to (9) yields:

$$R = R_0(y - T_0 r)^{n/b} \tag{10}$$

where R_o = constant.

To evaluate the constant, assume a linear city where a constant width is allocated to residentical location. Assume households are free to locate on either side of the *CBD*. Then equating supply and demand for land at any distance from the *CBD*:

$$2w_i dr = s[pf(r)]dr \tag{11}$$

where w_i = Width of city's land allocated to residential location
p = City's population
$f(r)$ = Measure assigning households to distance from the *CBD* in equilibrium

Substituting (7) and (10) into (11), rearranging and imposing the condition that all households be included, we require:

$$1 \equiv \int_{r=0}^{y/T_o} f(r)dr = \frac{2w_i R_o}{p\left(\frac{b}{n}\right)} \int_{r=0}^{y/T_o} (y - T_o r)^{(n/b)-1} dr \tag{12}$$

For convenience, it is assumed in (12) that the *CBD* occupies no space and that residential occupancy starts at the center of the city and extends to the point where all income is spent on commuting, i.e., $rT_o = y$. Alternatively, the city could be bounded at some minimum desirable level of population density or where land rent for residential occupancy falls to its value in agricultural production.

Performing the integration in equation (12) yields the residential land rent adjacent to the *CBD*, R_o. Substituting back into (10) the equilibrium land rent is found:

$$(13)\qquad R(r) = \left(\frac{pT_o}{2w_i}\right)\left(1 - \frac{T_o r}{y}\right)^{n/b}$$

Making further substitutions into (11) and (7), the equilibrium allocation of households to distance from the *CBD* and their pattern of land occupancy are also determined.

$$(14)\qquad f(r) = \left(\frac{nT_o}{by}\right)\left(1 - \frac{T_o r}{y}\right)^{(n/b)-1}$$

$$(15)\qquad s(r) = \left(\frac{2w_i yb}{pT_o n}\right)\left(1 - \frac{T_o r}{y}\right)^{1-(n/b)}$$

For later use when considering local government behavior, citywide expenditures on land, the composite other good, and on commuting are aggregated and shown as equations (16), (17), and (18).

(16)

$$sR_{tot} = \int_0^{y/T_o} \underbrace{\left[\left(\frac{b}{n}\right)(y - T_o r)\right]}_{\text{land expenditure per household}} \cdot \underbrace{[pf(r)]dr}_{\text{number of households}}$$

$$= \left(\frac{b}{n+b}\right)py$$

$$(17)\qquad xP_{tot} = \left(\frac{n-b}{n+b}\right)py$$

$$(18)\qquad T_o r_{tot} = \left(\frac{b}{n+b}\right)py$$

Notice that aggregate commuting and land rent expenditures are invariant with respect to the per mile commuting cost. As this cost rises people will crowd in toward the *CBD*, bidding up the land rent there so that although the individual location and expenditure patterns will be different, aggregate expenditures will not. Under this model's assumptions, provision of lower cost commuting will not adversely affect aggregate metropolitan area-wide land values.

By substituting (15) and (6) into (5), per household utility in a spatial equilibrium is derived in (19).

(19) Utility/household =

$$\underbrace{U_1[Q_oAW_{tot}]^c}_{U_o}\left(\frac{a}{nP}\right)^a\left(\frac{2w_i b}{npT_o}\right)^b y^n$$

As would be expected, this equilibrium level of utility is invariant with respect to the household's location. Everyone experiences the same level of utility. Notice also that increased commuting costs do adversely affect utility.

B. *Optimum Government Behavior*

At last the originally posed problem can be addressed. Assume a single, metropolitan area-wide government is trying to maximize aggregate welfare. One might dispute the realism of this optimizing criteria; nevertheless, it is interesting to examine the implied governmental policy.

First the revenue sources must be specified. A land value tax will be considered because of its desirable efficiency consequences. The supply of land in the city is perfectly inelastic in the sense that it is fixed at any specified distance from the *CBD*. Should land value tax revenues prove inadequate, consider an income tax as the alternative source.[12]

Next consider how these taxes affect the spatial equilibrium formulations of the preceding section. In particular, if it is assumed that each household owns the land which it occupies and that the land was originally obtained free of charge from the government, then the new household budget constraint is:

$$(20)\qquad y(1 - t_y) = xP + t_l sV(r) + T_o r$$

where t_y = Income tax rate
t_l = Land value tax rate
$V(r)$ = Value of land per unit area

[12] In the equal income community, an income tax is identical to a head tax.

Notice that although land may be purchased without charge, there is still a cost associated with owning it, namely the land value tax. Since the spatial equilibrium analysis is an investigation of long-run behavior, consider the land to be valued as a perpetuity.

$$(21)\qquad sV(r) = \frac{sR - t_l[sV(r)]}{i} \quad \text{or}$$

$$sV(r) = \frac{sR}{i + t_l}$$

where i = rate of return on all capital.

Then the budget constraint in (20) can be restated:

$$(22)\qquad y^* = xP + sR^*(r) + T_o r$$

where $y^* = (1 - t_y)y$

$$R^* = \left(\frac{t_l}{i + t_l}\right) R$$

Thus, a budget constraint of the form in (5) is still valid. With the exception that y has been replaced by y^* and R has been replaced by R^*, all of the spatial equilibrium results in Section IIIA still hold.

The variable W_{tot} and constant A, both of which are viewed as constants in the individual's utility function, must also be specified. Assume that there are no returns to scale in producing these public services[13] and that a Cobb-Douglas form may be used:

$$(23)\qquad W_{tot} = W_o K^\alpha L^\delta \qquad (\alpha + \delta = 1)$$

where K = Capital used by public sector
L = Labor used by public sector
W_0, α, δ = Constants

The constant A must satisfy the following condition:

$$(24)\qquad W_{tot} = \int_{r=0}^{y/T_o} \underbrace{[W_{tot} A \phi^g]}_{\substack{\text{public services}\\ \text{per household}}} \cdot \underbrace{[pf(r)]dr}_{\substack{\text{number of}\\ \text{households}}}$$

where the city is bounded by the point where all income is spent on commuting. Recalling that $\phi = 1/s$, substituting (14) and (15) in (24), integrating and rearranging, it can be shown that in order to maintain a spatial equilibrium:[14]

$$(25)\qquad A = \left[\frac{1 + g(1 - 2c)}{p(1 - cg)}\right] \cdot \left\{\frac{c(1 - g)}{(1 - cg)}\left[\frac{2w_i y}{pT_o}\right]\right\}^g$$

Aggregate utility levels are now completely specified. Government's budget constraint can be formulated as follows:

$$(26)\qquad TE = \omega L + \rho K = t_l s V_{tot} + t_y p y$$

where TE = total expenditures on public services
ω = the wage rate
ρ = the rental rate on capital $= iP_k$
P_k = the price of a unit of capital

Making use of (21), (22), and (16), the total revenue collected from land value taxes can be determined:

$$(27)\qquad t_l s V_{tot} = t_l\left[\frac{sR_{tot}}{i + t_l}\right] = sR^*_{tot} = \left(\frac{b}{n + b}\right) py^*$$

So long as local government selects and maintains a single land value tax rate and the interest rate remains constant, total receipts from the land value tax equal what the aggregate land rent would have been

[13] Werner Hirsch has summarized studies, most of which suggest that constant returns to scale are evident for a variety of local government services.

[14] Since public service levels cannot be negative at any location, $A \geq 0$. For $0 \leq c \leq 1$, the range of g is restricted to $-1 \leq g \leq 1$.

without a tax![15] Furthermore, government collects the same revenue from the land value tax regardless of the tax rate it establishes, and this revenue is the same as if government had retained ownership of the land and collected the land rent.[16]

Since unemployment is not considered in this model, the existence of a constant wage rate in (26) confronting a sizable employer such as local government presumes that the long-run private demand for labor is perfectly elastic. George Borts and Jerome Stein have presented a plausible set of circumstances which would produce this labor market condition.[17] Both private and public sectors are assumed to require the same labor skills, and workers not employed by local government will be snapped up by private industry.

Substitute (27) into (26) and specify government expenditures as some multiple q of aggregate land value tax collections (total land rent without a land value tax):

$$(28)\qquad TE = sR^*_{tot} + t_y py = qsR^*_{tot} = q\left(\frac{b}{n+b}\right)py^*$$

In (28) there is a direct relationship between the income tax rate t_y and q. Subsequent manipulations will be facilitated by considering q rather than t_y. Furthermore, q is of interest because if it is greater than one, an income tax is required, if less than one, an income subsidy is being provided (the land rent is returned to households) and where q equals one, the land rent just equals the optimal expenditures on government services. The remaining problem is to express y^*, the net income after taxes, as a function of known parameters. Multiplying the definition of y^* in (22) by population and using (28):

$$(29)\qquad py^* = py - t_y py = Y + (1-q)sR^*_{tot}$$

where $Y = py$ = aggregate wage and non-land investment income earned by local residents

But the spatial equilibrium value of sR^*_{tot} was given in (27) and the constants b and n were expressed in (5). Rearranging (29) and solving for y^*:

$$(30)\qquad y^* = \frac{Y}{p}\left[\frac{1+c(1-2g)}{(1-cg)+qc(1-g)}\right]$$

Aggregate utility is determined by multiplying (19) by the city's total population. The expression for W_{tot} in (23) and A in (25) are substituted. The expanded values for b and n from (5) are included, and y is replaced by y^* as determined in (30). Local government's budget constraint is presented in (28), and the welfare optimizing local government would seek to maximize the Langrangian expression shown in equation (31).

Local government now has four variables to select: K and L, the amounts of capital and labor to use in producing public services; g which determines how those services are to be distributed spatially; and q which determines the optimum level of expenditures on public services. The usual differentiation technique can be employed to solve for optimum values of K, L, and q as functions of g, but differentiating (31) with respect to g results in an unmanage-

[15] A similar result is obtained under the alternative assumption that all residential land is privately owned by its occupants but that it was originally purchased from the government. Here government collects land value taxes plus a return on the funds originally obtained from households in exchange for permitting private land ownership. Again government collects the equivalent, no-land-value-tax land rent regardless of the tax rate.

[16] Separate, interesting dynamic problems that might be examined are to determine the income distribution which would result from varying the tax rate after the original land sale or if local government initially sold all of the land at a price below a farsighted market value, and the city's population increases. In both cases, the spatial equilibrium analysis is more complex because the income distribution is endogenously determined.

[17] Essentially, it requires at least one local employer with a linear homogeneous production function whose product is sold competitively in a national market.

$$(31)\quad H = \frac{U_1\left[Q_oW_o\left(\frac{2w_ic}{T_o}\right)\right]^c}{p^{2c}}\left(\frac{1-c}{P}\right)^{1-c}\left\{\frac{(1-g)[1+g(1-2c)]}{(1-cg)}\right\}^c\left(\frac{Y}{1-cg}\right)$$
$$\cdot\left[\frac{1+c(1-2g)}{(1-cg)+qc(1-g)}\right]\cdot K^{\alpha c}L^{\delta c}+\lambda\left\{qY\left[\frac{c(1-g)}{(1-cg)+qc(1-g)}\right]-\omega L-\rho K\right\}$$

able expression. The first three steps result in equations (32) and (33).

$$(32)\quad K = \left(\frac{\alpha}{\rho}\right)\left(\frac{c}{1+c}\right)Y$$
$$L = \left(\frac{\delta}{\omega}\right)\left(\frac{c}{1+c}\right)Y$$
$$(33)\quad q = \frac{1-cg}{1-g}$$

The optimum value for g was explored by substituting (32) and (33) back into (31) and performing numerical analysis. The substitutions yield:

$$(34)\quad U_{tot} = \frac{U_c}{(1+c)^{1+c}}\left\{\frac{1+c(1-2g)}{(1-cg)^2}\right\}\cdot\left\{\frac{(1-g)[1+g(1-2c)]}{(1-cg)}\right\}^c$$

where

$$U_c = (1/p^{2c})\,U_1\left[Q_oW_o\left(\frac{2w_ic^2}{T_o}\right)\left(\frac{\alpha}{\rho}\right)^\alpha\left(\frac{\delta}{\omega}\right)^\delta\right]^c\cdot\left(\frac{1-c}{P}\right)^{1-c}Y^{1+c}$$

We can conveniently let $U_c=1$ and assign some arbitrary value to c where $0\leq c\leq 1$. Values of total utility in (34) can then be simulated for a range of g where it was previously shown $-1<g<1$. The results are sketched in Figure 1.

A welfare maximum is attained where public services are slightly skewed in favor of more densely population areas ($g>0$) thus encouraging denser configurations. The benefit is a decrease in aggregate commuting expenditures which are a loss to society. Those inhabitants who live in sparsely populated areas are compensated for the loss of public services by experiencing even lower levels of population density and land rent. Furthermore, the utility maximizing distribution of public services is Pareto optimal for this community with equal incomes since in equilibrium, every household must have the same level of utility.

However, if public services are allocated too heavily in favor of densely inhabited areas, aggregate welfare begins to fall. Density rises more rapidly than public services under a fixed resource constraint. In the limit as g approaches one, the ever increasing proportion of public services in dense areas just offsets the underlying negative externalities, and everyone will

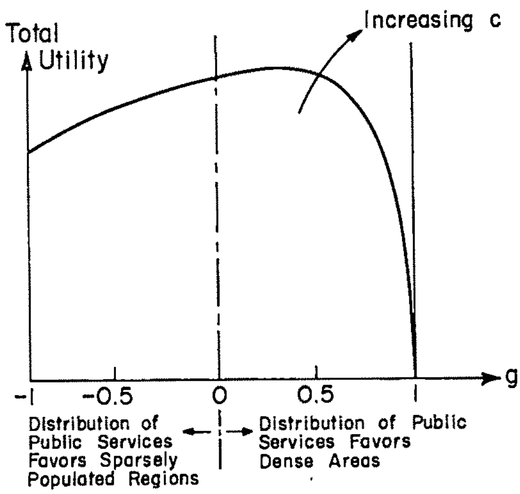


FIGURE 1. CITY WITH EQUAL PER HOUSEHOLD INCOMES

choose to live at the *CBD* to avoid commuting costs. But in the limit as density at the *CBD* approaches infinity, no finite level of public services can offset the underlying externality, and aggregate welfare will approach zero.

The optimum income tax rate can be determined by rearranging (28) and substituting the defined relationship between y and y^* from (22) and the optimum value for q from (33). This yields:

$$t_y = \left(\frac{1-c}{1+c}\right)\left(\frac{cg}{1-cg}\right) \tag{35}$$

Optimum values of t_y from (35) are plotted in Figure 2, and the resultant levels of aggregate land rent and total government expenditures are shown in Figure 3 as functions of g. Total land rent is obtained by substituting (33) in (30) and this in turn into (27).

$$sR^*_{tot} = \left[\frac{c(1-g)}{(1+c)(1-cg)}\right] Y \tag{36}$$

Total government expenditures are found by substituting (32) into (26) (or (36) and (33) into (28)).

$$TE = \left(\frac{c}{1+c}\right) Y \tag{37}$$

Notice in Figure 3 that the level of optimum expenditures on public services is independent of their distribution. Where

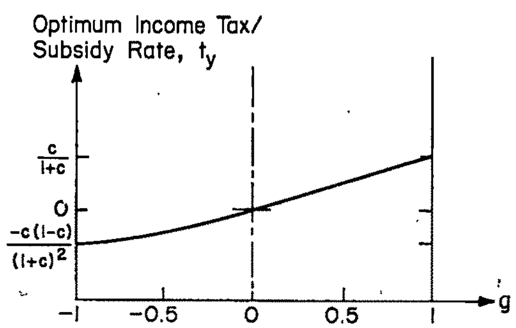


FIGURE 2

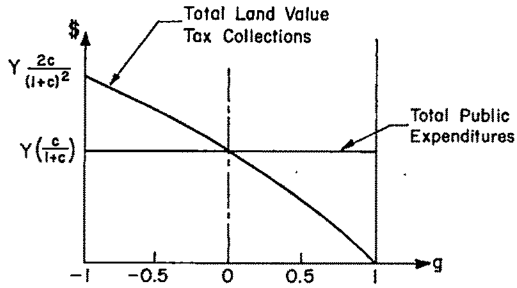


FIGURE 3

$g>0$, a welfare maximizing local government would seek to raise additional revenue from the income tax to augment the declining land rent. Conversely, if local government were to skew the distribution of its services to favor sparsely settled areas, land rent collections would exceed the optimal public expenditures, and local government should strive to transfer the excess back to households.

Clearly, a government policy of avoiding income taxes is suboptimal unless they also follow the suboptimal policy of allocating public services equally to each household.

IV. A City with Varying Household Incomes

It is interesting to examine how optimum government policies change when the realism of the model is improved by allowing household incomes to vary. Assume each household's tastes remain identical, but that incomes vary with some known distribution. The first-order conditions of Section II will still hold, but the solution for a spatial equilibrium will be different. Unfortunately (9) can no longer be solved directly for the equilibrium rent gradient since income will now be some function of distance from the *CBD*. In place of (11), write:

$$2w_i dr = sp \underbrace{\left[f(y)\left|\frac{dy}{dr}\right|\right]}_{f(r)} dr \tag{38}$$

Since a one-to-one relationship is required between income and distance, $f(r)$, the distribution of households as a function of distance may be represented as a transformation of $f(y)$, the distribution of households as a function of their income. Here $|dy/dr|$ is the Jacobian of the transformation. But for a household in equilibrium, $dr/dy>0$, and the absolute value signs may be dropped. Substituting (7) in (38) and rearranging:

$$(39) \qquad R = \left[\frac{\left(\frac{b}{n}\right)p}{2w_i}\right](y - T_o r)f(y)\frac{dy}{dr}$$

Taking the logarithm of (39), differentiating with respect to distance and equating the result to (9) yields the following second-order, non-linear differential equation relating income level to distance:

$$(40) \qquad \frac{d\log\left[(y - T_o r)f(y)\frac{dy}{dr}\right]}{dr} = \frac{-T_o\left(\frac{n}{b}\right)}{(y - T_o r)}$$

If (8) is substituted into (38) and rearranged, the result is a relationship between land rent and income.

$$(41) \qquad \frac{dR}{dr} = \frac{-pT_o}{2w_i}f(y)\frac{dy}{dr}$$

In order to proceed, assume incomes are exponentially distributed.

$$(42) \qquad f(y) = \frac{e^{-y/B}}{B} \qquad (0 \leq y \leq \infty)$$

where $E(y)=B$

If, as in the city with equal incomes, it is assumed that land rent falls to zero at the outer reaches of the city, consistent closed form solutions which satisfy both (40) and (41) may be obtained:

$$(43) \qquad r = \frac{y - B\left(\frac{n}{b}\right)[1 - e^{-(b/n)y/B}]}{T_o}$$

$$(44) \qquad R = \left(\frac{pT_o}{2w_i}\right)e^{-y/B}$$

However, here the city is unbounded since people with infinite incomes can live as far from the *CBD* as they like. The equilibrium consumption of space and other goods by each household are:

$$(45) \qquad s = \left(\frac{2w_iB}{pT_o}\right)[1 - e^{-(b/n)y/B}]e^{y/B}$$

$$(46) \qquad x = \left(\frac{aB}{bP}\right)[1 - e^{-(b/n)y/B}]$$

As incomes become large, households will live very far from the *CBD* and occupy enormous amounts of space, but their consumption of other goods will approach the constant (aB/bP).

Again it is interesting and a necessary input for subsequent analysis to examine aggregate expenditure patterns.

$$(47) \qquad sR_{tot} = \int_{y=0}^{\infty}\underbrace{\left[\left(\frac{b}{n}\right)(y - T_o r)\right]}_{\substack{\text{land expenditures}\\ \text{per household}}}\cdot\underbrace{[pf(y)]dy}_{\substack{\text{number of}\\ \text{households}}}$$

$$= \left(\frac{b}{n+b}\right)pB$$

Similarly:

$$(48) \qquad xP_{tot} = \left(\frac{n-b}{n+b}\right)pB$$

$$(49) \qquad T_o r_{tot} = \left(\frac{b}{n+b}\right)pB$$

As in the city with equal incomes, a constant fraction of aggregate incomes, pB, is spent on each good and on commuting regardless of the unit commuting costs. Individual expenditure patterns will vary but the aggregates are constant.

In this case per household utility will obviously vary with location since it will vary with income level, and a social welfare

function must be introduced in order to consider optimal government policy. Martin Feldstein's assumption of a power function will be used to weight individual utility levels.

$$(50) \qquad SW = U^m \qquad (0 \leq m \leq 1)$$

where SW = social welfare
m = elasticity of welfare due to changes in utility

The range specified for m is arbitrary, but it guarantees constant or diminishing marginal welfare. The lower limit represents equal weighting of all households regardless of income level, whereas the upper limit implies a direct additive weighting of utilities. When the spatial equilibrium results from (45) and (46) are included together with the known income distribution in (42), government's objective function becomes:

$$(51) \quad SW_{tot} = \int_{y=0}^{\infty} [U_o x^a s^b]^m [pf(y)] dy$$
$$= \left[U_1 (Q_o A W_{tot})^c \left(\frac{a}{bP}\right)^a \cdot \left(\frac{2w_i}{pT_o}\right)^b B^n \right]^m p \left(\frac{n}{b}\right) \beta_1$$

where β denotes the *beta* function and $\beta_1 = \beta\{(nm+1); n[(1/b)-m]\}$.

As in the city where all households have equal incomes, if local government supplements land value tax revenues with a proportional income tax, all of the spatial equilibrium results in (43) through (51) still hold if y is replaced by y^*, net income after taxes, and the average income B is replaced by B^*, the average after tax income. Furthermore, the equations for government's budget constraint (26) through (30), are still valid if, in this case, y is replaced by B and y^* is replaced by B^* since per household income in the equal income city is identical to average income.

These substitutions presume that both government and industry hire the various skill levels represented by different incomes in the same, invariant proportions. The only factor substitution is between capital and labor, not among the various classes of labor.[18]

The final alteration is to evaluate A for inclusion in (51) through U_o. The procedure is the same as in (24) and the result can be shown to be:

$$(52) \quad A = \frac{\left[\frac{c(1-g)}{p(1-cg)}\right]\left(\frac{2w_i B^*}{pT_o}\right)^g}{\beta_2}$$

$$\text{where } \beta_2 = \beta\left[(1-g); \frac{(1-cg)(1+g)}{c(1-g)}\right]$$

By again making all the necessary substitutions from (5) and (52) into (51) and using (28) as the budget constraint (the same as in (31)), a Lagrangian expression can be developed. When maximized with respect to K, L, q, and λ, results identical to those in (32), (33), and (35) are obtained for the optimum use of labor and capital and the optimum income tax. However, when (32) and (33) are substituted back into the expression for social welfare in (51), equation (53) is obtained, where U_c was defined in (34).

$$(53) \quad SW_{tot} = U_c^m p^{1-m} \left[\frac{1-cg}{c(1-g)}\right] \cdot \left\{\frac{1+c(1-2g)}{c^{1-c}(1-g)^{1-2c}[(1-cg)(1+c)]^{1+c}}\right\}^m \left\{\frac{\beta_1}{\beta_2^{cm}}\right\}$$

By letting $U_c^m p^{1-m} = 1$, and assuming specific values for c,[19] the level of social wel-

[18] Another implicit assumption of this spatial equilibrium analysis is that physical relocation is far more rapid than alterations in workers' skills.

[19] In the example of Figure 4, the assumption of $c=.1$ implies that approximately 10 percent of disposable income in urban areas is spent on land rent.

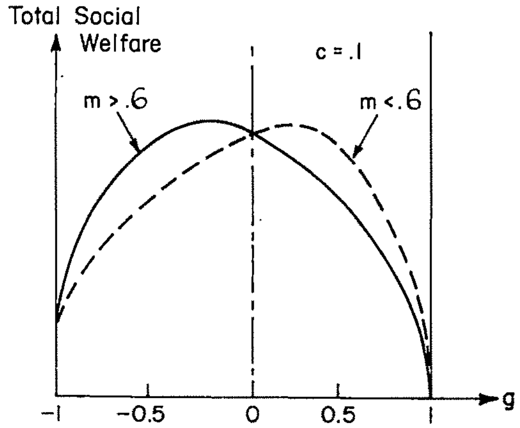


FIGURE 4. CITY WITH EXPONENTIALLY DISTRIBUTED HOUSEHOLD INCOME

fare can be analyzed as a function of g for various values of m. The results are shown in Figure 4.

In this case, the optimal policy for allocating public services hinges upon the parameter value in society's welfare function. For the numerical example in Figure 4, the crucial value is where a 10 percent increase in utility represents a 6 percent change in social welfare. For a more egalitarian community ($m<.6$), social welfare is maximized if the public service allocation is weighted in favor of densely inhabited areas. Here the results and the reasons are similar to those for the city where all households have the same income. Favoring densely populated areas encourages dense settlement patterns and it discourages commuting. In addition, since with $g>0$ land value taxes will provide insufficient revenue to finance the optimal amount of public services, a supplemental income tax will be required. The proportional income tax is used to finance public services which in this case are distributed disproportionately in favor of densely populated areas where the poor live. This result is not surprising under a social welfare function which places decreasing marginal value on increasing levels of individual utility. Here it dictates a public policy that shifts increments of utility from high to low income households through the provision and financing of public services.

However, when $m>.6$, even over a range where diminishing marginal social welfare is exhibited, the direction of optimal government policy is seen to reverse in Figure 4. Here, the price theorist's prescription for dealing with externalities is borne out, although only after outweighing two contradicting tendencies. Favoring sparsely settled areas through the allocation of public services ($g<0$) encourages all households to occupy more land thereby reducing population density throughout the city.[20] This in turn reduces the need for public services. On the negative side, households seeking more land will move further from the *CBD* and incur larger commuting costs which represents a deadweight loss to society. Also, where ($g<0$), Figure 3 shows that land value taxes are more than adequate to finance the optimal level of public services. The excess revenues are returned as a subsidy that is proportional to income. Both the allocation of services and the subsidy favor higher income households—policies which are not encouraged by according diminishing social welfare to individual utility levels.

The basic difference between the city with exponentially distributed incomes and the equal income community that allows these negative effects to be overcome is that the city with varying incomes is very much larger. The small proportion of households with infinite incomes can always move further from the *CBD* and occupy more space. Thus when the public service distribution is skewed to favor sparsely settled areas, *all* households can occupy more space. By comparison, the

[20] This can be seen by examining the derivative of (45). See the author.

city in which all households have the same income is bounded by commuting expenditures, and attempts to occupy more land by those living close to the *CBD* imposes higher densities on those living close to the city's boundaries.

V. Summary and Policy Comparisons

Local government's services in urban areas have been conceptually divided into three categories: those that facilitate accessability (transportation); wealth redistribution and enhancement (welfare and education); and services which offset externalities which arise from dense habitation (police and fire protection, sanitation, etc.). Focusing on the last group, it was shown that because of possible household relocation, there are many interactions between service level, residential location patterns, and government revenue available through land value taxes. Therefore the standard economic prescription for handling negative externalities—that of taxing their generation or subsidizing alternatives—does not always apply within this complex framework. This has been shown using a spatial equilibrium model of urban form. Here, depending upon the degree of concavity of the community's welfare function, or in a city where all households have the same income, it may be optimal to allocate density-offsetting public services in favor of densely populated neighborhoods which in turn encourages dense habitation patterns. This tendency arises in the *CBD* focused city because it decreases commuting costs which represent a social loss.

However, if the welfare function is only mildly concave in a community where household incomes are exponentially distributed, the traditional economic policy of discouraging patterns of dense habitation which generate negative externalities is optimal. Here the public service distribution is skewed in favor of sparsely settled areas. The resultant overall settlement patterns are less dense and contribute more to welfare than is deducted through increased commuting. This reversal of optimal policy casts suspicion on prescriptions derived from models where all households have equal incomes. The basic formulation of a spatial equilibrium changes when some known distribution of incomes is assumed; although the assumption of a specific welfare function is required in order to investigate public policy.

This analysis does suggest that a government which follows an optimal policy of allocating public services in favor of dense areas is doomed to financial difficulty unless it supplements land value taxes with some other revenue source (see Figure 3). Conversely, if an opposite allocation policy is followed, the government will enjoy a revenue bonanza in the sense that land value tax receipts exceed expenditures. Unfortunately, such a policy may be suboptimal depending on the community's welfare function. Finally, in order for an equal per capita allocation of local public services to be optimal requires a unique social welfare function. At the very least, Henry George's advocates should be dismayed to discover that in many cases, land value taxes should prove inadequate to finance one category out of the several which comprise the total bundle of local public services.

Many other interesting analyses can be made within this spatial equilibrium framework. Taxes other than those on income may be investigated as supplements to land value taxes. (See the author.) A production function for commuting could be included so that tradeoffs between local government's expenditures to facilitate commuting and on density off-setting services could be analyzed. A further complication would be to include population density in these production functions. (Solow has included space on the com-

muting side.) Other possible investigations include government organization, air pollution patterns under multiple sources, and the effect of education.

However, the policy relevant question at this stage is whether a static, long-run equilibrium model which allows consideration of many simultaneous interactions sufficiently captures the essence of urban problems, or are their origins more closely linked to difficulties in adjusting to change? The first step toward an answer would require empirical testing to see if urban dynamics are sufficiently slow to allow mature areas to approach an equilibrium.

REFERENCES

W. Alonso, *Location and Land Use*, Cambridge 1964.

W. Baumol, "On Taxation and the Control of Externalities," *Amer. Econ. Rev.*, June 1972, *62*, 307–22.

M. Beckmann, "On the Distribution of Urban Rent and Residential Density," *J. Econ. Theory*, June 1970, *1*, 60–67.

G. Borts and J. Stein, *Economic Growth in a Free Market*, New York 1964.

J. Buchanan, "An Economic Theory of Clubs," *Economica*, Feb. 1965, *32*, 1–14.

M. Feldstein, "Distributional Equity and the Optimal Structure of Public Prices," *Amer. Econ. Rev.*, Mar. 1972, *62*, 32–36.

W. Hirsch, "The Supply of Urban Services," in H. Perloff and L. Wingo, eds., *Issues in Urban Economics*, Baltimore 1968, 477–525.

H. Hoyt, "Recent Distortions of the Classical Models of Urban Structure," *Land Econ.*, May 1964, *40*, 199–212.

K. Lancaster, "A New Approach to Consumer Theory," *J. Polit. Econ.*, Apr. 1966, *74*, 132–58.

J. Lansing and E. Mueller, *The Geographic Mobility of Labor*, Ann Arbor 1967.

P. Mieszkowski, "Tax Incidence Theory: The Effects of Taxes on the Distribution of Income," *J. Econ. Lit.*, Dec. 1969, *7*, 1103–24.

E. Mills, *Studies in the Structure of the Urban Economy*, Baltimore 1972.

A. Montesano, "A Restatement of Beckmann's Model on the Distribution of Urban Rent and Residential Density," *J. Econ. Theory*, Apr. 1972, *4*, 329–54.

R. Muth, *Cities and Housing*, Chicago 1969.

W. Oates, "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis," *J. Polit. Econ.*, Nov./Dec. 1969, *77*, 957–71.

J. Ohls and T. Wales, "Supply and Demand for State and Local Services," *Rev. Econ. Statist.*, Nov. 1972, *54*, 424–30.

A. M. Polinsky and S. Shavell, "Air Pollution and Property Values in a General Equilibrium Model of an Urban Area," Urban Inst. working pap. no. 1207–5, Washington July 1972.

P. A. Samuelson, "The Pure Theory of Public Expenditures," *Rev. Econ. Statist.*, Nov. 1954, *36*, 387–89.

R. E. Schuler, "The Interaction Between Local Government and Residential Location in an Urban Area," unpublished doctoral dissertation, Brown Univ. 1972.

R. M. Solow, "Congestion, Density and the Use of Land in Transportation," *Swedish J. Econ.*, Mar. 1972, *74*, 161–73.

C. M. Tiebout, "A Pure Theory of Local Expenditures," *J. Polit. Econ.*, Oct. 1956, *64*, 416–24.

L. Wingo, Jr., *Transportation and Urban Land*, Washington 1961.

A Note on the Elasticity of Derived Demand Under Decreasing Returns

By Assaf Razin*

In his classic book, *The Theory of Wages*, John Hicks analyzes Marshall's four rules for the factors on which the elasticity of derived demand depends.[1] The smaller the elasticity, the stronger the trade union power may be expected to be in the industry. Recently, Ryuzo Sato and Tetsunori Koizumi (1970) have generalized the Marshall-Hicks analysis into situations of multiplicity of factors of production, allowing for various possibilities of substitution and complementarity in production. While these studies were done in the context of a perfectly competitive industry, Patrick Yeung developed rules of derived demand in the context of imperfect competition. The traditional analysis is confined, however, to the assumption of "constant returns." This assumption may be inappropriate for short-run analyses where some factors of production are temporarily immobile. If one is interested, therefore, in the short-run properties of the demand for factors of production, the assumption of "decreasing returns" may be more plausible. The purpose of this note is to derive the general formula and to discuss the properties of a derived demand for a factor of production under conditions of decreasing returns to scale.

I. The Model

Consider a perfectly competitive industry in which a product X is being made by the cooperation of two factors L and K under conditions of decreasing returns to scale. If the elasticity of demand for the product (in absolute values) is η and the elasticity of supply of K is e, how is λ, the elasticity of demand for L (in absolute values), determined?

To analyze this problem it is convenient to transform the decreasing-returns-to-scale production function $X=F(L, K)$ into a constant-returns-to-scale function $G(L, K, T)=TF(L/T, K/T)$, by adding a fictitious factor T. This transformation was first introduced in the 1973 paper by Eitan Berglas and the author. If we set $T=1$, profits will be equal to the (shadow) price of T, P_T. The prices of L, K are denoted by P_L, P_K, respectively; the price of X is P.

II. The Elasticity of Derived Demand

Consider the set of equilibrium conditions for a perfectly competitive industry which faces a given demand schedule for its product $X(P)$ and a given supply schedule of a factor K, $K(P_K)$:

$$(1)\qquad P_La_L + P_Ka_K + P_Ta_T = P$$

$$(2)\qquad a_LX = L$$

$$(3)\qquad a_KX = K$$

$$(4)\qquad a_TX = 1 \qquad (\text{or } T = 1)$$

where a_i is the cost-minimizing (for given prices P_L, P_K, P_T) input-output coefficient of factor i; a_i is a function of factor prices

$$(5)\quad a_i = a_i(P_L, P_K, P_T), \qquad i = L, K, T$$

Equation (1) has cost per unit of output plus profits per unit of output on the left-hand side and the price of the product on the right-hand side. Equations (2) and (3) describe full employment of factors L and K while equation (4) describes equality of demand and supply for the fictitious factor T.

The elasticities of the input-output coefficients can be written in terms of R. G. D.

* Department of economics, Tel-Aviv University. The paper was written in spring 1973, while I was a visiting professor at the University of Minnesota. I wish to thank, without implicating, John Hause and Tetsunori Koizumi for useful comments on an earlier draft.

[1] A more explicit mathematical statement of the problem is found in Martin Bronfenbrenner, Appendix to ch. 6.

$$(7)\quad \begin{bmatrix} -1 & \alpha_L & \alpha_K & \alpha_T \\ -\eta & \alpha_L\sigma_{LL}+\lambda & \alpha_K\sigma_{LK} & \alpha_T\sigma_{LL} \\ -\eta & \alpha_L\sigma_{KL} & \alpha_K\sigma_{KK}-e & \alpha_T\sigma_{KT} \\ -\eta & \alpha_L\sigma_{TL} & \alpha_K\sigma_{TK} & \alpha_T\sigma_{TT} \end{bmatrix} \begin{bmatrix} d\log P \\ d\log P_L \\ d\log P_K \\ d\log P_T \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Allen's partial elasticities of substitution σ_{ij}:[2]

$$(6)\quad \frac{P_j}{a_i}\frac{\partial a_i}{\partial P_j} = \alpha_j\sigma_{ij}, \qquad \sum_{j=L,K,T} \alpha_j = 1,$$

$$\sum_{j=L,K,T} \alpha_j\sigma_{ij} = 0, \qquad i = L, K, T$$

By taking the total differential and making use of the definition of the various elasticities and conditions for cost minimization ($\sum P_i da_i=0$, $i=L, K, T$), the set of equations (1)–(4) can be converted into the elasticity form in equation (7). The vanishing of the determinant of the coefficient matrix in (7), required for nontrivial solutions, provides the formula for λ in an implicit form as equation (8).

$$(8)\quad \det \begin{bmatrix} \frac{1}{\eta} & 1 & 1 & 1 \\ 1 & \sigma_{LL}+\frac{\lambda}{\alpha_L} & \sigma_{LK} & \sigma_{LT} \\ 1 & \sigma_{KL} & \sigma_{KK}-\frac{e}{\alpha_K} & \sigma_{KT} \\ 1 & \sigma_{TL} & \sigma_{TK} & \sigma_{TT} \end{bmatrix} = 0$$

Hicks' formula, derived for the case of constant returns to scale, can be obtained as a special case of (8) in the following way. Consider the case where the production function is *homogeneous* of degree $1-\alpha_T$. Then, it is easy to see that since

$$(1/T)^{1-\alpha_T}F(K, L) = F(K/T, L/T)$$

the transformation $G=TF(K/T, L/T)$ takes a simple form, $G=T^{\alpha_T}F(K, L)$.[3] The partial elasticities of substitution of T and other factors reduce in this case to

$$(9)\quad \sigma_{TT} = -\frac{1-\alpha_T}{\alpha_T}, \qquad \sigma_{Ti} = 1 \qquad \text{for } i = L, K$$

Substitute (9) into (8), multiply the last column by α_T, and let α_T approach zero (homogeneity of degree one) to get (10).

$$(10)\quad \det \begin{bmatrix} \frac{1}{\eta} & 1 & 1 & 0 \\ 1 & \sigma_{LL}+\frac{\lambda}{\alpha_L} & \sigma_{LK} & 0 \\ 1 & \sigma_{KL} & \sigma_{KK}-\frac{e}{\alpha_K} & 0 \\ 1 & \sigma_{TL} & \sigma_{TK} & -1 \end{bmatrix} = 0$$

Expansion of (10), using $\sigma_{LL}=-(\alpha_K/\alpha_L)\sigma$, $\sigma_{KK}=-(\alpha_L/\alpha_K)\sigma$, $\sigma_{LK}=\sigma_{KL}=\sigma$ yields

$$(11)\quad \lambda = \frac{\eta\sigma + e(\alpha_L\eta + \alpha_K\sigma)}{\alpha_L\sigma + \alpha_K\eta + e}$$

which is Hicks' formula, p. 244.

Returning to the general case, we can establish rules for elasticities of derived demand using the novel procedure employed in Sato and Koizumi (1970). Applying the method of differentiation between elements in the determinant in (8), we get

$$(12)\quad \frac{\partial\lambda}{\partial\eta} = \frac{\alpha_L}{\eta^2}\frac{\Delta_{00}}{\Delta_{11}}$$

$$(13)\quad \frac{\partial\lambda}{\partial e} = \frac{\alpha_L}{\alpha_K}\frac{\Delta_{22}}{\Delta_{11}}$$

[2] See Allen, p. 504. For a detailed discussion of the relationships among different concepts of elasticities of substitution which were used in the literature, the reader may consult Sato and Koizumi (1973).

[3] See the 1972 paper by Berglas and the author for general equilibrium analysis in this case.

where Δ_{ij} is the cofactor of the element in the ith row and jth column in the determinant in (8), $(i, j=0, 1, 2)$. By a theorem on zero-valued and symmetric determinants (see Thomas Muir, ch. 7), we have $\Delta_{ii}\Delta_{jj}=\Delta_{ij}^2$. This together with (12)–(13) implies that the more elastic is the demand for the product or the supply of the cooperant factor, the more elastic will be the derived demand for L. These are, respectively, Marshall's "fourth" and "third" rules of derived demand.[4] The effect on λ of a change in σ_{KL}, compensated by changes in σ_{LL} and σ_{KK} (satisfying (6)), is given by

$$(14)\qquad \frac{\partial\lambda}{\partial\sigma_{KL}}=\frac{1}{\alpha_K\Delta_{11}}\left[\alpha_L^2\Delta_{11}-2\alpha_L\alpha_K\Delta_{12}+\alpha_K^2\Delta_{22}\right]$$

Manipulating the right-hand side of (14) and using $\Delta_{11}\Delta_{22}=\Delta_{12}^2$, we get

$$(14a)\qquad \frac{\partial\lambda}{\partial\sigma_{KL}}=\frac{1}{\alpha_K}\left(\frac{\Delta_{22}}{\Delta_{11}}\cdot\alpha_L-\alpha_K\frac{\Delta_{12}}{\Delta_{22}}\right)^2>0$$

This result is a generalization for the case of decreasing returns of Marshall's "first rule." As already shown by Hicks, Marshall's "second rule" with regard to the effect of the magnitude of the factor's distributive share (in total revenue) on the elasticity of derived demand for the factor cannot be determined a priori, even in the case of constant returns to scale. Under decreasing returns to scale, this ambiguity also prevails.

III. The Degree of Decreasing Returns

An economic interpretation of the fictitious factor's own elasticity of substitution can now be given. Treat parametrically the price of the product and consider the elasticity of supply of X, $\epsilon=(\partial X/\partial P)(P/X)$. This can be obtained by totally differentiating (1) and (4), using the cost minimization condition $\sum P_i da_i=0$, and converting into elasticities, to get (15).

$$(15)\qquad \begin{pmatrix}\alpha_T & 0\\ \alpha_T\sigma_{TT} & 1\end{pmatrix}\begin{pmatrix}d\log P_T\\ d\log X\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}d\log P-\begin{pmatrix}\alpha_L\\ \alpha_L\sigma_{TL}\end{pmatrix}d\log P_L-\begin{pmatrix}\alpha_K\\ \alpha_K\sigma_{TK}\end{pmatrix}d\log P_K$$

Setting $d\log P_L=d\log P_K=0$ and solving (15) with respect to $\epsilon=\partial\log X/\partial\log P$ yields

$$(16)\qquad \epsilon=-\sigma_{TT}$$

Therefore, the fictitious factor's own elasticity of substitution is equal in absolute values to the elasticity of supply of the product.[5]

Suppose we measure the *degree of decreasing returns* by α_T.[6] What is the effect of decreasing returns on the elasticity of derived demand? To give an answer to this question, we differentiate between elements in the determinant in (8) where the change in α_T is compensated for by changes in α_K, σ_{TT}, and σ_{KK} as follows:

$$(17)\qquad \frac{\partial\lambda}{\partial\alpha_T}=\frac{\alpha_L}{\Delta_{11}}\left[\frac{\alpha_L\sigma_{KL}+(1-\alpha_L)\sigma_{KT}+e}{\alpha_K^2}\Delta_{22}+\frac{\sigma_{TK}-\sigma_{TT}}{\alpha_T}\Delta_{33}\right]$$

Therefore, if factors K and L are substitutes in production ($\sigma_{KL}>0$) and if K is a normal input ($\sigma_{TK}-\sigma_{TT}>0$),[7] the smaller

[4] There is, however, an exceptional case of a neutral input. A neutral input is an input whose quantity does not change when the product price increases while factors' prices are held constant. Let L be a neutral input. By a well-known proposition in the theory of production (see C. E. Ferguson, ch. 9), $\partial L/\partial P=-\partial X/\partial P_L$. In this case, a change in P_L will not directly affect production. Therefore, when L is a neutral input and $e=\infty$ (so that P_K is unchanged), λ is independent of η.

[5] Note that in the case of constant returns to scale $-\sigma_{TT}=\infty$.

[6] Note (from (1)) that $1-\alpha_T$ is the reciprocal of elasticity of total costs (which is equal to the ratio of marginal and average costs).

[7] A normal (inferior) input is an input whose quantity increases (decreases) when the price of the product increases while factors' prices are kept constant. Differentiating the left-hand side of (3), holding P_L and P_K constant, and using the solutions for $d\log P_T/d\log P$ and $d\log X/d\log P$ from (15), we get: $\partial\log K/\partial\log P=\sigma_{TK}-\sigma_{TT}$. Therefore, K is a normal or an inferior input if $\sigma_{TK}-\sigma_{TT}$ is greater than or smaller than zero, respectively. For a discussion of inferior inputs, see Moses Syrquin.

the degree of returns to scale α_T, the more elastic is the derived demand λ. However, when K and L are complements in production ($\sigma_{LK}<0$), the sign of $\partial\lambda/\partial\alpha_T$ cannot be determined a priori. Also, when α_K is held constant, the results of the effect of a change in α_T compensated for by changes in α_L, σ_{LL}, and σ_{TT} are ambiguous in the case of complementarity in production.

IV. Distributive Share

It may be of interest to analyze, for the case of decreasing returns, the effect of a change in the price of the factor on the distributive share of this factor. Consider the distributive share of the factor L:

$$\alpha_L = \frac{a_L P_L}{P} \tag{18}$$

We consider, as did Hicks (in ch. 6), the case of an infinitely elastic demand for the product and an infinitely elastic supply of the cooperant factor. Differentiating the right-hand side of (18), substituting the solutions for $d\ log\ X/d\ log\ P_L$ and $d\ log\ P_T/d\ log\ P_L$ from (15) (when $d\ log\ P=d\ log\ P_K=0$), we get:

$$\frac{P_L}{\alpha_L}\frac{\partial\alpha_L}{\partial P_L} = \alpha_K(1-\sigma_{LK}) + (1-\alpha_K)(1-\sigma_{LT}) \tag{19}$$

The first term on the right-hand side of (19) is due to the substitution of K for L for a given level of output, while the second is due to the effect on the distributive share of the change in output. To verify that the latter is an expansion effect, we analyze the effect of a change in the product price on α_L. Differentiating (18) while holding P_L and P_K constant and using the solutions for $d\ log\ X/d\ log\ P$ and $d\ log\ P_T/d\ log\ P$ from (15), we get

$$\frac{P}{\alpha_L}\frac{\partial\alpha_L}{\partial P} = \sigma_{LT} - 1 \tag{20}$$

When the production function is *homogeneous* (i.e., $\sigma_{LT}=1$), equation (19) reduces to

$$\frac{P_L}{\alpha_L}\frac{\partial\alpha_L}{\partial P_L} = \alpha_K(1-\sigma_{LK}) \Rightarrow \frac{\partial\alpha_L}{\partial P_L} \gtreqless 0 \quad \text{as } 1 \gtreqless \sigma_{LK} \tag{19a}$$

Equation (19a) is a generalization of Hicks' condition for the change in the factor's distributive share, when the price of the factor changes. Obviously, if the production function is of a Cobb-Douglas form, $\sigma_{LK}=\sigma_{LT}=1$, we get $\partial\alpha_L/\partial P_L=0$.

REFERENCES

R. G. D. Allen, *Mathematical Analysis for Economists*, London 1956.

E. Berglas and A. Razin, "Protection and Real Profits," working pap. no. 14, Foerder Inst., Tel-Aviv Univ., Dec. 1972.

——— and ———, "Effective Protection and Decreasing Returns to Scale," *Amer. Econ. Rev.*, Sept. 1973, *63*, 733–37.

M. Bronfenbrenner, *Income Distribution Theory*, Chicago 1972.

C. E. Ferguson, *The Neoclassical Theory of Production and Distribution*, Cambridge 1969.

J. R. Hicks, *The Theory of Wages*, New York 1932.

T. Muir, *A Treatise on the Theory of Determinants*, Rev. by W. H. Metzler, New York 1960.

R. Sato and T. Koizumi, "Substitutability, Complementarity, and the Theory of Derived Demand," *Rev. Econ. Stud.*, Jan. 1970, *37*, 107–18.

——— and ———, "The Production Function and the Theory of Distributive Shares," *Amer. Econ. Rev.*, June 1973, *63*, 484–89.

M. Syrquin, "A Note on Inferior Inputs," *Rev. Econ. Stud.*, Oct. 1970, *37*, 591–98.

P. Yeung, "A Note on the Rules of Derived Demand," *Quart. J. Econ.*, Aug. 1972, *86*, 511–17.

A Note on the Structure of Optimal Taxation

By AGNAR SANDMO*

The recent articles by William Baumol and David Bradford (1970) and by Avinash Dixit would seem together to form the perfect introduction to modern theories of optimal taxation. The former gives a broad view of the historical and institutional background for the analysis and derives the main results from a series of simplified models. The latter sets up an explicit second best optimization model and proves the main theorems in a rigorous manner. Thus, the pair of articles should serve well as a basic survey of the field and as a preparation for the more general formulations of Peter Diamond and James Mirrlees, and Joseph Stiglitz and Partha Dasgupta.

Unfortunately, however, the careful reader will soon be struck by an apparent inconsistency between the two analyses. While Baumol and Bradford derive the classic rule that tax rates should be inversely proportional to price elasticities (in the case of zero cross-elasticities), Dixit concludes differently. On what actually seems to be a more general set of assumptions he claims to prove that a proportional tax structure, i.e., a structure in which the tax *rates* are the same for all goods, is in fact optimal, and that deviations from this rule must be sought in the presence of untaxable goods. I shall argue here that there is an inner contradiction in Dixit's analysis of the case where all goods are taxable, and that the correct version of his analysis yields conclusions that are in fact consistent with those of Baumol and Bradford.

The contradiction referred to above has also been noted by Yew-Kwang Ng, but the exchange between him and Bradford and Baumol (1972) did not quite clear up the matter. Since the issue is an important one, a more detailed reconsideration of it seems to be in order. Sections I–III present essentially a reformulation of the Dixit model and derive results similar to those of Baumol and Bradford. This exercise is deemed worthwhile because the Dixit model represents a more direct approach to the second best optimization problem involved. Section IV analyzes some further aspects of the question of when a proportional (uniform) tax is optimal.

I

I shall adopt the simplifying assumptions that preferences can be represented by a social utility function, that producer prices are given, and that the public sector is concerned to raise a given amount of tax revenue. The objective of the public sector is to maximize the utility of consumers subject to the tax revenue constraint. This is probably the simplest model one can work with, and it does of course have the shortcoming that distributive effects of alternative tax schemes are ignored. The assumption of given producer prices is on the other hand less restrictive than it may seem, since it is shown both by Diamond and Mirrlees and by Dixit that the results carry over to the case of variable producer prices if there are constant returns to scale.

The social utility function is given by

$$(1) \qquad U = (x_0, x_1, \ldots, x_m)$$

which is taken to have the usual concavity properties. Consumers face prices $P_i = p_i + t_i$, where p_i are producer prices and t_i are taxes. Factors of production are treated as negative consumption goods; $t_i > 0$ for $x_i < 0$ then implies that consumers are subsidized on

* Norwegian School of Economics and Business Administration. I am indebted to Victor Norman for many discussions and to Camille Bronsard, Avinash Dixit, Takashi Negishi, and a referee for helpful comments and criticism.

their supply of factor i. The budget constraint of consumers is written by Dixit as

$$\sum_{i=0}^{m} P_i x_i = I \tag{2a}$$

where I is defined as "transfer payments, if any." But to assume $I>0$ is quite a paradox in a model whose basic assumption is that transfer payments are not allowed.[1] This is in fact pointed out by Bradford and Baumol (1971), who suggest that this accounts for the differences with their own analysis; this point will be taken up later. We shall here make the assumption (not ruled out by Dixit) that $I=0$, so that the budget constraint becomes

$$\sum_{i=0}^{m} P_i x_i = 0 \tag{2b}$$

This condition is in fact a statement of Walras' law within the context of this model.

The public sector's tax revenue constraint can be written as

$$\sum_{i=0}^{m} t_i x_i = T \tag{3}$$

Define the tax rates θ_i as t_i/P_i. Equation (3) can then be rewritten as

$$\sum_{i=0}^{m} \theta_i P_i x_i = T \tag{3'}$$

Suppose that all goods are taxed at the same rate θ. Then we have that

$$\sum_{i=0}^{m} \theta_i P_i x_i = \theta \sum_{i=0}^{m} P_i x_i = 0 \tag{4}$$

where the last equality follows from (2b). In an economy where Walras' law holds, a proportional tax on all goods will yield zero tax revenue. The intuitive interpretation of this is of course that proportional taxation requires either that all consumer goods are taxed and all factors are subsidized at the same rate, or that consumer goods are subsidized and factors are taxed at the same rate. In either case total subsidy payments will exhaust tax revenue. Dixit's result that proportional taxation is optimal is only meaningful when it is assumed that $I>0$.

II

In the following it will be assumed that commodity 0 is not taxed and is used as numéraire, so that $P_0=p_0=1$ and $t_0=0$. A natural interpretation of this might be that commodity 0 is labor and that the other commodities are consumer goods; this makes the model similar to that of Baumol and Bradford. But it should be stressed that this does involve some loss of generality, since we have ruled out the case where all goods are taxed, but at different rates. Put differently, we limit the analysis to an investigation of the optimal structure of indirect taxation rather than of commodity taxation in general, including taxes on labor.

The first-order conditions for utility maximization by consumers, given the budget constraint (1), are

$$U_i - \lambda P_i = 0 \qquad i = 0, 1, \ldots, m \tag{5}$$

Here λ is the Lagrange multiplier which can be interpreted as the marginal utility of income. The demand functions are

$$x_i = x_i(P) \qquad i = 0, 1, \ldots, m \tag{6}$$

where P is the price vector $(P_1, \ldots, P_m)$. The indirect utility function is then defined as

$$V(P) = U(x(P)) \tag{7}$$

and it is easy to show, utilizing (2b) and (5), that

$$\frac{\partial V}{\partial P_k} = -\lambda x_k \qquad k = 1, \ldots, m \tag{8}$$

[1] This statement refers to the *net* transfer to the private sector. $I=0$ does not preclude transfers within the private sector, but the existence of such transfers is immaterial for the following argument.

Note that, since producer prices are given, we must have $\partial V/\partial P_k = \partial V/\partial t_k$.

We can now formulate the maximization problem of the public sector in terms of the Lagrangian function

$$L = V(P) - \mu\left(\sum_{i=1}^{m} t_i x_i - T\right) \tag{9}$$

Setting the partial derivatives of this expression with respect to t_k equal to zero, we have that

$$\frac{\partial V}{\partial P_k} - \mu\left(\sum_{i=1}^{m} t_i \frac{\partial x_i}{\partial P_k} + x_k\right) = 0 \qquad k = 1, \ldots, m \tag{10}$$

Utilizing (8), we can rewrite the conditions as

$$-\lambda x_k = \mu\left(\sum_{i=1}^{m} t_i \frac{\partial x_i}{\partial P_k} + x_k\right) \qquad k = 1, \ldots, m \tag{11}$$

$$\sum_{i=1}^{m} t_i \frac{\partial x_i}{\partial P_k} = -\frac{\lambda + \mu}{\mu} x_k = -\nu x_k \qquad k = 1, \ldots, m \tag{12}$$

Differentiating (2b) with respect to P_k we obtain

$$\frac{\partial x_0}{\partial P_k} + \sum_{i=1}^{m} P_i \frac{\partial x_i}{\partial P_k} + x_k = 0 \tag{13}$$

Substituting for x_k in (12) yields

$$\sum_{i=1}^{m} \theta_i P_i \frac{\partial x_i}{\partial P_k} = \nu\left(\sum_{i=1}^{m} P_i \frac{\partial x_i}{\partial P_k} + \frac{\partial x_0}{\partial P_k}\right) \tag{14}$$

which can be rewritten as

$$\sum_{i=1}^{m} P_i(\theta_i - \nu) \frac{\partial x_i}{\partial P_k} = \nu \frac{\partial x_0}{\partial P_k} \qquad k = 1, \ldots, m \tag{15}$$

This corresponds to equations (8) in Dixit's article. The difference lies in that he has the right-hand sides all equal to zero, since all goods are taxed. He then concludes that a proportional tax with $\theta_i = \nu$ is optimal, but this is meaningless in a world where Walras' law holds.

In general, we cannot conclude from (15) that proportional (uniform) taxes are optimal. However, it is clearly of interest to look for special cases in which the solution to (15) does in fact imply proportionality. One such case is clearly that where $\partial x_0/\partial P_k = 0$ for all k, so that the demand for leisure (supply of labor) is completely inelastic with respect to all consumer goods prices. A classic insight of taxation theory is that goods, which are inelastic in demand, are ideal tax objects from the efficiency point of view. In this case we have barred ourselves from taxing labor by choosing it as the numéraire, but by levying a proportional tax on all consumer goods, we do in fact achieve the same result. The point is that we are changing only the relative price of commodity 0, while within the group of taxed goods all relative prices are unchanged.

We can reinterpret the Dixit model so as to make it consistent with the present analysis by assuming that the "transfer income" is actually a fixed stock of commodity 0 which functions as a numéraire. This interpretation is logically consistent, but we must remind the reader of the very strong conditions needed for such a fixed resource to exist.[2]

III

We now turn to the inverse elasticity formula. Assume that

$$\frac{\partial x_i}{\partial P_k} = 0 \quad \text{for } i \neq k \qquad i, k = 1, \ldots, m \tag{16}$$

This means that, *within the group of taxed*

[2] The discussion of this problem by Leif Johansen, ch. 7, seems to rest on a similar assumption. He assumes the existence of exogenous "income," and in the transformation function the only variables are the quantities of consumer goods. Together these assumptions must imply that labor is a fixed resource.

goods, all cross-derivatives of the demand functions are zero. From (12) it then follows that

$$\theta_k = -\frac{\nu}{\epsilon_k} \qquad k = 1, \ldots, m. \tag{17}$$

where $\epsilon_k \equiv (\partial x_k/\partial P_k)(P_k/x_k)$. This is the inverse elasticity formula. Dixit argues that the standard implication that taxes should be levied primarily on goods that are inelastic in demand is not really valid, since with demand independence all price elasticities must be equal to minus one. This does not follow here, however, since no assumption is made about the derivatives $\partial x_0/\partial P_k$. If, in addition to the independence assumption, one assumes $\partial x_0/\partial P_k = 0$ for all k, we are back to a proportional structure, but this is simply a special case of the more general result derived in the previous section.

IV

One of the most valuable features of Dixit's article is his distinction between taxable and nontaxable goods, and that part of his analysis is not affected by the arguments presented here. On the contrary, the present analysis confirms them. When there exist nontaxable goods, Dixit concludes that if tax increases do not lead to any shifting of demand to nontaxable goods, proportional taxes are optimal. This is obviously equivalent to our condition that price changes do not lead to any change in the demand for leisure. Moreover, Dixit's conclusion that the inverse elasticity formula is valid if cross-derivatives of demand functions vanish within the group of taxable goods is borne out by the results of the previous section of this paper.

To sum up: Taxing all commodities at the same rate is meaningless in a general equilibrium model where Walras' law holds. Choosing labor as an untaxed numéraire good, we have shown how the inverse elasticity formula may validly be derived from the Dixit model, and we have shown that proportional taxation is optimal in the special case where the numéraire good is in completely inelastic demand.

The latter conclusion does not, however, represent a *necessary* condition for proportional taxation to be optimal, and the question naturally arises whether there are other conditions of general economic interest which lead to the same conclusion. This problem has been studied in an article by Anthony Atkinson and Stiglitz, and there is no reason to repeat their analysis here. Since their analysis is somewhat different from the present one, it may, however, be worthwhile to see how the problem may be approached from a different angle.

From (13) we have that if

$$\frac{\partial x_0}{\partial P_k} = \alpha x_k \qquad k = 1, \ldots, m \tag{18}$$

where α is some function which is independent of k, then a proportional tax will be optimal. For (13) then becomes

$$\sum_{i=1}^{m} P_i \frac{\partial x_i}{\partial P_k} = -(1+\alpha)x_k \tag{19}$$

and after substitution of x_k from (12) we obtain

$$\sum_{i=1}^{m} P_i \left(\theta_i - \frac{\nu}{1+\alpha}\right)\frac{\partial x_i}{\partial P_k} = 0 \qquad k = 1, \ldots, m \tag{20}$$

This corresponds to (15), and we see that a proportional tax structure provides a solution to these equations. Moreover, if the matrix $(\partial x_i/\partial P_k)$ is nonsingular, the solution will be unique. Note that our previous condition is in fact the special case where $\alpha = 0$.

Expanding the left-hand side of (18) in terms of the Slutsky equation we obtain

$$\frac{\partial x_0}{\partial P_k} = -x_k\frac{\partial x_0}{\partial I} + S_{0k}, \qquad k = 1, \ldots, m \tag{21}$$

where S_{0k} are the substitution terms and $\partial x_0/\partial I$ is the income effect, which is here evaluated at $I = 0$. We see immediately that

since the income effect is in fact proporportional to x_k, one condition for (18) to be satisfied is that the cross-substitution terms vanish. This means that proportional taxation will be optimal if there are no relationships of complementarity or substitutability between labor (or leisure) and consumer goods.

Evidently, we can be a bit more general than this, since it suffices for this result that we can write

$$(22) \qquad S_{0k} = \beta x_k \qquad k = 1, \ldots, m$$

where β is again some function independent of k. For (21) can then be rewritten as

$$(23) \qquad \frac{\partial x_0}{\partial P_k} = \left(- \frac{\partial x_0}{\partial I} + \beta \right) x_k = \alpha x_k \qquad k = 1, \ldots, m$$

Equation (22) can be interpreted as the requirement that there be no *special* relationship of complementarity or substitutability between labor (leisure) and consumer goods. But this is a very loose statement. Is there a particular class of preference orderings which will yield conditions (22) on the substitution terms?

Let us assume that the utility function is weakly separable, in the sense defined by Steven Goldman and Hirofumi Uzawa, between labor and consumption goods. It can then be written as

$$(24) \qquad U = U(x_0, F(x_1, \ldots, x_m))$$

By a straightforward adaptation of Theorem 5 of Goldman and Uzawa we then have that the substitution effects satisfy the condition

$$(25) \qquad S_{0k} = \kappa \frac{\partial x_0}{\partial I} \cdot \frac{\partial x_k}{\partial I} \qquad k = 1, \ldots, m$$

where κ is some function of the arguments of the utility function U. We can rewrite (25) as

$$(26) \qquad S_{0k} = \kappa I^{-2} x_0 E_0 E_k x_k \qquad k = 1, \ldots, m$$

where E_i is the income or Engel elasticity of commodity i. We now make the additional assumption that the function F is homogeneous (of arbitrary positive degree), representing a homothetic indifference map for consumer goods. This means that the proportions in which two commodities are consumed depends only on their relative prices and not on income; hence they must have the same income elasticities, i.e., $E_j = E_k$ for all j, $k = 1, \ldots, m$. But then we can define

$$(27) \qquad \beta(x_0, x_1, \ldots, x_m) = \kappa I^{-2} x_0 E_0 E_k$$

and thus (26) implies (22).[3] We have then proved that if the utility function is weakly separable between labor and consumption goods and homogeneous in consumption goods, then proportional taxation is optimal.

This result represents an alternative derivation of a result of Atkinson and Stiglitz. It should be observed that it can be generalized immediately to the case discussed by Dixit where there are several nontaxable goods.

V

We have seen that there are some interesting special cases in which proportional taxation will be optimal. Still, these cases are special enough to justify the conclusion that proportionality will be the exception rather than the rule if an optimal tax structure is to be chosen on the basis of pure efficiency criteria. Empirical explorations of optimal tax structures will be valuable contributions to further study of this problem.[4]

[3] The fact that we have previously assumed $I = 0$ is irrelevant for this argument, where the income elasticities are simply used to characterize properties of the indifference map.

[4] Atkinson and Stiglitz take an important step towards implementation of the theory. An ambitious attempt at empirical application is contained in an unpublished paper by Camille Bronsard et al.

REFERENCES

A. B. Atkinson and J. E. Stiglitz, "The Structure of Indirect Taxation and Economic Efficiency," *J. Publ. Econ.*, Apr. 1972, *1*, 97–119.

W. J. Baumol and D. F. Bradford, "Optimal

Departures from Marginal Cost Pricing," *Amer. Econ. Rev.*, June 1970, *60*, 265–83.

D. F. Bradford and W. J. Baumol, "Optimal Taxes and Pricing: Reply," *Amer. Econ. Rev.*, Mar. 1972, *62*, 175–76.

C. Bronsard et al., "La Pratique du Second Best: L'Optimation des Taxes Indirectes Canadiennes," disc. pap., Univ. Montréal 1973.

P. A. Diamond and J. A. Mirrlees, "Optimal Taxation and Public Production: I," *Amer. Econ. Rev.*, Mar. 1971, *61*, 8–27.

——— and ———, "Optimal Taxation and Public Production: II," *Amer. Econ. Rev.*, June 1971, *61*, 261–78.

A. K. Dixit, "On the Optimum Structure of Commodity Taxes," *Amer. Econ. Rev.*, June 1970, *60*, 295–301.

S. M. Goldman and H. Uzawa, "A Note on Separability in Demand Analysis," *Econometrica*, July 1964, *32*, 387–98.

L. Johansen, *Public Economics*, Amsterdam 1965.

Y-K. Ng, "Optimal Taxes and Pricing: Comment," *Amer. Econ. Rev.*, Mar. 1972, *62*, 173–74.

J. E. Stiglitz and P. Dasgupta, "Differential Taxation, Public Goods and Economic Efficiency," *Rev. Econ. Stud.*, Apr. 1971, *38*, 151–74.

The Diagrammatic Analysis of Multiple Consumption Externalities

By EDGAR K. BROWNING*

Most analyses of consumption externalities have been based on the assumption that only one of the goods consumed by an individual involves external effects.[1] Clearly, however, many cases exist where the consumption of more than one good generates external effects. This paper develops a diagrammatic technique capable of analyzing situations when these multiple consumption externalities are present. More specifically, we will assume that an individual consumes two goods both of which affect the welfare of others, and then show how the optimal allocations of resources can be represented diagrammatically. This technique will then be used to compare the efficiency of lump sum transfers and consumption subsidies when multiple external benefits in consumption exist.[2]

Increasingly in recent years economists have based arguments for public subsidization of education, medical care, housing, food, and other goods, squarely on the external benefits associated with their consumption. Judging from these arguments, we are certainly in a world characterized by multiple consumption externalities, yet the argument for a public subsidy is based on the (usually implicit) assumption that there is only one consumption externality present. Since the consumption of food, housing, and medical care probably account for three-fourths of the expenditures of many families, especially poor families (to whom the argument for a subsidy is often restricted), we are clearly more accurate in assuming that all goods consumed (at least by the poor) generate external benefits than we are in basing the analysis on the assumption that only one good does. When we formulate the analysis on the more realistic assumption, some rather surprising conclusions emerge. For instance, we will show that when external benefits are associated with all goods consumed by an individual, there is no valid a priori case on grounds of efficiency which can be made for a public subsidy of any kind.

I

We will assume that society is composed of two individuals, A and B. There are three goods produced, X, Y, and Z. Individual B consumes only X and Y, in quantities X_B and Y_B; his utility function can be expressed as:

$$U_B = U_B(X_B, Y_B) \tag{1}$$

Individual A consumes only good Z (which may, of course, be a Hicksian composite good composed of X, Y, and other goods). In addition, A's utility depends on B's consumption of both X and Y so A's utility function can be written as

$$U_A = U_A(X_B, Y_B, Z_A) \tag{2}$$

It will be further assumed that all partial derivatives (marginal utilities) are positive. Then the interdependence exhibited in equations (1) and (2) describes a situation of nonreciprocal external benefits in consumption (i.e., A's welfare depends on B's consumption, but not vice versa). However, the geometrical technique which is described below is quite flexible, and can be adapted to the analysis of reciprocal externalities, either positive or negative.

* Assistant professor of economics, University of Virginia. I wish to thank Edgar Olsen, Roger Sherman, and the managing editor for helpful suggestions at several stages in the preparation of this paper. Research support was provided by a Wilson Gee Fellowship at the University of Virginia for the summer of 1972.

[1] Perhaps the best geometric technique available for analyzing this case was developed by F. Trenery Dolbear, Jr.

[2] George Daly and Fred Giertz (1972a, b) have analyzed multiple consumption externalities in the context of a pure exchange economy. The model developed in this paper differs primarily in the incorporation of production possibilities into the analysis.

Production possibilities for the society are given by the linear constraint:

$$(3) \qquad P_xX + P_yY + P_zZ = K$$

where K is a constant and P_x, P_y, and P_z are the constant marginal costs, or prices, of X, Y, and Z, respectively. The production possibilities for the three goods are shown in Figure 1 as the transformation plane $T_1T_2T_3$. Any point on the transformation plane implies a specified level of consumption for both A and B since the goods have been defined in such a way that B consumes all the X and Y produced, and A consumes all the Z. Thus, any point on the transformation plane uniquely determines the utilities of both A and B, and our main objective in the next several paragraphs will be to show how the utility levels of A and B can be represented in a two-dimensional graph.

Let us consider first how to represent A's rankings for the points on the transformation plane. Since A's utility depends on three variables, X_B, Y_B, and Z_A, his preferences can be shown as a set of three-dimensional indifference surfaces. These indifference surfaces are not drawn in, but perhaps can be most easily visualized by borrowing a description from William Vickrey and thinking of the three dimensional indifference surfaces as "... looking something like a nest of bowls tilted against the corner of a cupboard, the amounts of the three commodities represented by the distances from a given point to the two walls and the floor ..." (p. 34). In Figure 1, the "bowls" would be tilted against the origin, with those lying further from the origin representing successively higher levels of welfare. Thus, the highest indifference surface attainable would be the bowl farthest from the origin given the constraint implied by the transformation plane. This indifference surface would be tangent to the $T_1T_2T_3$ plane at only one point, which is shown as point E in Figure 1. The allocation of resources at point $E(X_B, Y_B, Z_A)$ is the allocation that maximizes A's welfare.[3]

[3] The allocation that maximizes A's welfare may be a corner solution at T_1, of course.

Lower levels of utility for A occur where the indifference surfaces intersect the transformation plane. Thus, a bowl lying closer to the origin will have its sides slicing the transformation plane in a circular shape (it need not be a perfect circle, of course), showing all the combinations of X, Y, and Z on the transformation plane which imply this lower level of utility. Therefore, lower levels of utility will be shown as circular shapes lying farther from point E on the transformation plane. For example, A will prefer point C to D, but D and K will be equally preferred.

In this way we can represent A's level of utility corresponding to any allocation of resources by the circular curves on the transformation plane. A's preferences, however, may also be shown in the two-dimensional X-Y space. A two-dimensional treatment is not only easier to handle, but, properly

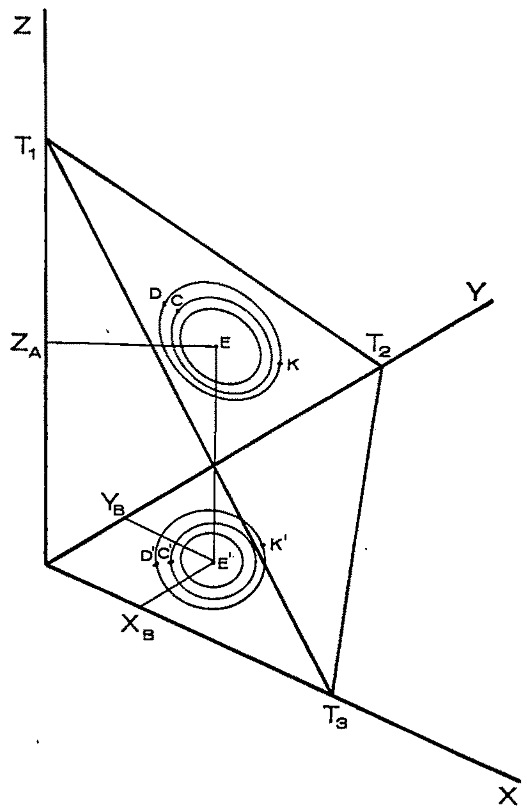


FIGURE 1

interpreted, also contains all the relevant information of the three-dimensional treatment. Since any point on the $T_1T_2T_3$ plane represents a combination of X and Y, the X-Y coordinates of the circular curves can be projected onto the X-Y plane. Point E' in the X-Y plane shows the best combination of X and Y(X_B and Y_B), given by E on the transformation plane. Similarly, D and K become D' and K' in the X-Y plane. The curves in the X-Y plane will, of course, retain a circular shape.

Figure 2 reproduces the curves in the X-Y plane in more detail. These curves will be called indifference contours to distinguish them from more orthodox indifference curves. The indifference contours reflect A's rankings of alternative combinations of three goods. For any given X_B and Y_B, the transformation plane specifies a unique quantity of Z_A so A's preferences can be shown as a function of only X_B and Y_B as long as production possibilities remain unchanged. As we move from one point to another on an indifference contour in Figure 2, the quantity of Z_A is also changing in a specified way (as determined by the transformation plane) to keep A at the same level of utility. Even though point K' in Figure 2 implies more of both X_B and Y_B than point D', individual A is indifferent between these allocations because he has less Z_A at point K'. (These points are shown in Figure 1 for additional clarity.)

Now let us consider the slope of an indifference contour in the X-Y plane. Rearranging the equation of production possibilities, the quantity of Z consumed by A can be expressed as:

$$(4) \quad Z_A=(K/P_z)-(P_x/P_z)X_B-(P_y/P_z)Y_B$$

Substituting (4) into the equation of one of A's indifference *surfaces* [$U_A(X_B, Y_B, Z_A)=a$ (a constant)], we have an equation for one of A's indifference *contours* expressed in terms of X and Y alone:

$$(5) \quad U_A[X_B,\ Y_B,\ (K/P_z)-(P_x/P_z)X_B - (P_y/P_z)Y_B]-a=0$$

Implicit differentiation of this expression

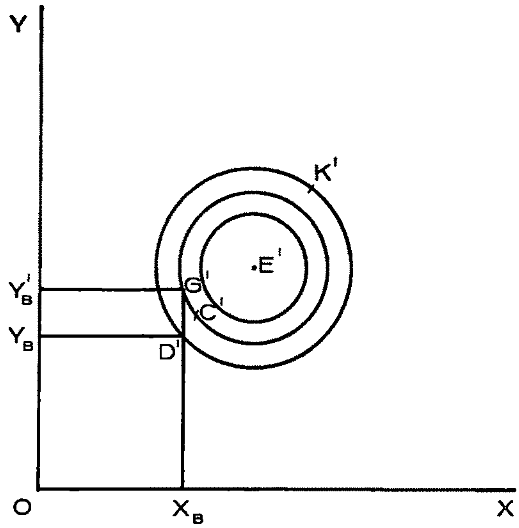


FIGURE 2

yields:

$$(6) \quad dy/dx=-\frac{\partial U_A/\partial X_B+\partial U_A/\partial Z_A(-P_x/P_z)}{\partial U_A/\partial Y_B+\partial U_A/\partial Z_A(-P_y/P_z)}$$

Dividing numerator and denominator by $-\partial U_A/\partial Z_A$, and writing $-\partial U_A/\partial X_B/\partial U_A/\partial Z_A$ as MRS^A_{xz} (A's marginal rate of substitution between X_B and Z_A, which is negative for an external benefit), we have the slope of an indifference contour:

$$(7) \qquad dy/dx = -\frac{P_x/P_z + MRS^A_{xz}}{P_y/P_z + MRS^A_{yz}}$$

Note that the slope of an indifference contour may be either positive or negative. For instance, an indifference contour will have a negative slope when the numerator and denominator on the right-hand side of (7) are both negative, i.e., when $-MRS^A_{xz}>P_x/P_z$ and $-MRS^A_{yz}>P_y/P_z$. This simply means that the marginal external benefit of each good is greater than its price. Point D' illustrates such a case. Note that higher indifference contours lie to the northeast. Thus, a move from D' to G' means that A is paying the entire cost of the increment in Y_B, but A will still be better off since $-MRS^A_{yz}>P_y/P_z$.

An indifference contour can also be

negatively sloped when the numerator and denominator of (7) are both positive. This occurs when $P_x/P_z > -MRS^A_{xz}$, and $P_y/P_z > -MRS^A_{yz}$. Thus, when the price of each good is greater than its marginal external benefit, the indifference contour will have a negative slope. In this case, however, higher indifference contours lie to the southwest, as at point K' in Figure 2. As we will see later, all efficient allocations lie in this region.[4]

Having represented A's preferences with indifference contours, we can easily determine A's level of welfare corresponding to any allocation of resources on the transformation plane. To determine whether or not a resource allocation is efficient, we must also know how B's welfare depends on the allocation of resources. Individual B's preferences can be represented with normally shaped indifference curves, as shown in Figure 3, since his utility depends only on his own consumption of X and Y. The curve T_2T_3 is simply the transformation curve relating X and Y when Z is equal to zero, from Figure 1 (with a slope equal to $-P_x/P_y$). Any point in Figure 3 shows a combination of X_B and Y_B, which determines B's level of welfare from equation (1), and also implies a level of Z_A, thereby determining A's level of welfare from equation (2). For instance, point K refers to the resource allocation X_B, Y_B, and $Z_A = KL$ times P_y/P_z.

The locus of all efficient allocations is shown by the contract curve, EC. Note that the contract curve is the locus of *some* of the tangencies between indifference curves of B and indifference contours of A. Thus a necessary condition for an efficient allocation of resources is:[5]

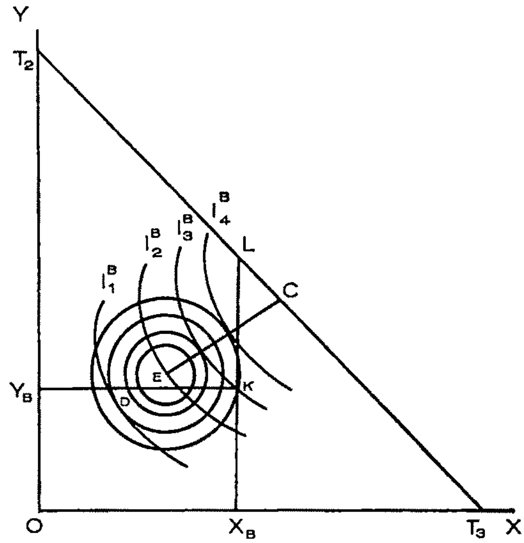


FIGURE 3

$$MRS^B_{xy} = -\frac{P_x/P_z + MRS^A_{xz}}{P_y/P_z + MRS^A_{yz}} \tag{8}$$

But not all points of tangency between B's indifference curves and A's indifference contours are efficient. The curves are tangent at D, but this is obviously inefficient since both parties would be better off at E. Thus, (8) is a necessary but not sufficient condition for efficiency.

Up to this point, nothing has been said about the efficiency or inefficiency of a competitive market equilibrium, or about the nature of policy to achieve efficiency when the market equilibrium is inefficient. In the following section the geometry developed here is adapted to analyze these issues, and some unusual results emerge. For example, it will be shown that a competitive equilibrium is not necessarily incompatible with an efficient allocation of resources in the presence of multiple consumption externalities. If, on the other hand, the competitive equilibrium is inefficient, our analysis indicates that a lump sum sub-

[4] In the region where the indifference contours are positively sloped, the marginal external benefit of one good is less than its price while the marginal external benefit of the other good is greater than its price.

[5] When there are fewer than two consumption externalities, this condition simplifies into a more familiar condition for efficiency. Thus, when there is no consumption externality, equation (8) becomes the familiar condition that B's marginal rate of substitution be equal to (minus) the price ratio. When there is only one externality (one $MRS^A = 0$), then the expression can be simplified to produce the familiar summed marginal rates of substitution equal to (minus) the price ratio.

sidy may be all that is required to attain efficiency.

II

A competitive equilibrium can be depicted by specifying the distribution of income between A and B. In Figure 4, B's budget constraint is shown as MN, with a slope equal to $-P_x/P_y$. Individual B's income, in units of Y, is equal to OM, and A's income, also in units of Y, is equal to $MT_2(=MT_2$ times P_y/P_z in units of $Z)$. Equilibrium occurs at K where B is maximizing his utility subject to his budget constraint, and A is consuming MT_2 times P_y/P_z units of Z. (In the following discussion of the competitive equilibrium and possible subsidy approaches to attaining efficiency, the possibility of a voluntary bargaining solution will be disregarded since our main concern is with the large numbers case in which bargaining is unlikely to occur.)[6]

Whether or not the competitive equilibrium represents an efficient allocation of resources depends on the precise shape of A's indifference contours. Figure 4 illustrates a case where the competitive equilibrium at K is efficient since it lies on the contract curve EC. Although there are external benefits associated with B's consumption of both goods, there is no move possible from the competitive equilibrium which will benefit one person without harming the other.

To see why such an outcome is possible, recall that the private equilibrium condition for B is:

$$(9) \qquad -P_x/P_y = MRS^B_{xy}$$

[6] When there is a large number of externally affected parties the indifference contours can be interpreted as social indifference contours with a slope of

$$-\frac{P_x/P_z + \sum_{i=1}^{n} MRS^i_{xz}}{P_y/P_z + \sum_{i=1}^{n} MRS^i_{yz}}$$

where the summation extends over all externally affected parties. In this case it is not necessary that every externally affected party receive external benefits from B's consumption of both X_B and Y_B for the conclusions of this paper to hold; some people may be affected only by X_B and others only by Y_B.

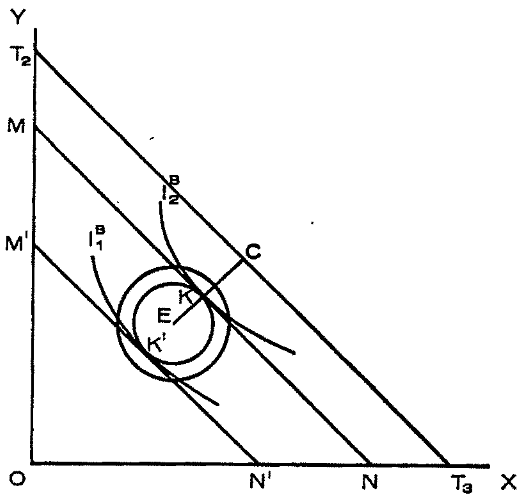


FIGURE 4

Recall equation (8) which gives the necessary condition for efficiency as MRS^B_{xy} equal to the slope of an indifference contour of A. Therefore, if the slope of an indifference contour of A equals $-P_x/P_y$ at B's private equilibrium, it will also equal MRS^B_{xy}, and a necessary condition for efficiency will be satisfied. For the competitive equilibrium to be efficient, then, we require that:

$$(10) \qquad P_x/P_y = \frac{P_x/P_z + MRS^A_{xz}}{P_y/P_z + MRS^A_{yz}}$$

Note that this equality will hold when $MRS^A_{xy}/P_y/P_z = MRS^A_{yz}/P_y/P_z$. This means that A is willing to pay the same fraction of the market price of X and Y to increase B's consumption of these goods, and nothing in a competitive system precludes this possibility. (In partial equilibrium terms this occurs when the marginal external benefit per dollar's worth of every good consumed by B is equal.) Therefore, a competitive equilibrium may satisfy the necessary condition for an efficient allocation of resources in the presence of external benefits in consumption.

The rationale underlying this conclusion is, of course, that we are dealing with a second best situation characterized by two distortions which affect the allocation of

resources in opposite, and offsetting, ways. An external benefit associated only with X leads B to consume *too much* Y, while an external benefit associated with Y alone leads him to consume *too little* Y (and the opposite for X). When the two distortions exist simultaneously, they tend to offset one another, and the net effect may be an efficient allocation of resources, as shown at K in Figure 4.

Competition, however, does not necessarily produce an efficient allocation of resources in the presence of external benefits associated with consumption of both X and Y. Consider the distribution of income (in Y units) of OM' for B and $M'T_2$ for A. Individual B chooses point K' on the $M'N'$ budget constraint. Although the slope of B's indifference curve and A's indifference contour are equal at K', this is not an efficient allocation since both A and B would be better off at points between E and K on the contract curve. A subsidy can be used to attain an efficient allocation, but note that a lump sum subsidy will be appropriate. A lump sum subsidy of MM' (and a lump sum tax of MM' for A) will shift B's budget constraint to MN, and B's new equilibrium at K will be an efficient one. Separate subsidies of B's consumption of both X and Y are not necessary to achieve efficiency because no change in relative prices is required.

In both cases, the competitive equilibrium involves a tangency between one of B's indifference curves and one of A's indifference contours, but in only one instance (at K) is the equilibrium an efficient allocation. At both K and K' we have $MRS^A_{xz}/P_x/P_z = MRS^A_{yz}/P_y/P_{yz}$. The difference lies in the value of these equal fractions at K and K'. At K', the fractions are less than minus one (i.e., greater than one in absolute value) indicating that A is willing to pay more than the market price of an increment in B's consumption of both X and Y. When this is true, efficiency obviously requires an increase in B's consumption of both goods. At K, on the other hand, both fractions, though equal, are greater than minus one, indicating that A is willing to pay only part of the market price of an increment in B's consumption of the goods. When this is the case, the competitive equilibrium is efficient, and there is no way to benefit B except by harming A since an increase in B's consumption of both X and Y requires that A pay the entire cost of the increments. Therefore, a competitive equilibrium is efficient when we have, at B's privately attained equilibrium:

$$\frac{MRS^A_{xz}}{P_x/P_z} = \frac{MRS^A_{yz}}{P_y/P_z} \geq -1 \tag{11}$$

When (11) holds there will be a tangency between B's indifference curve and A's indifference contour and with higher indifference contours located to the southwest.

When, at B's equilibrium, $MRS^A_{xz}/P_x/P_z \neq MRS^A_{yz}/P_y/P_z$, the slope of A's indifference contour will not be equal to P_x/P_y, and hence not equal to MRS^B_{xy}, so the competitive equilibrium will be inefficient. This will necessarily be true, regardless of the value of the two ratios. And in this situation, the familiar conclusion that a consumption subsidy for one of the two goods will attain efficiency is correct. It is only necessary to subsidize one of the goods despite external benefits being associated with both. Roughly speaking, a subsidy for the good with the larger marginal external benefit is all that is required. A case of this type is illustrated in Figure 5 where at the equilibrium K, $MRS^A_{xz}/P_x/P_z < MRS^A_{yz}/P_y/P_z$. Note that K does not lie on the contract curve EC of efficient resource allocations. For this particular configuration of preferences, one of the efficient allocations can be achieved with a per unit subsidy of the consumption of X which shifts B's budget constraint to MN'. Then B's equilibrium occurs at K' which lies on the contract curve. It is not surprising that a subsidy for X alone will work in this case since, in partial equilibrium terms, the marginal external benefit per dollar's worth of X is greater than the marginal external benefit per dollar's worth of Y. Thus, encouraging B to substitute X for Y is a move in the appropriate direction. No subsidy for Y is required despite the external benefit associated with its consumption. This is clearly another instance where reasoning

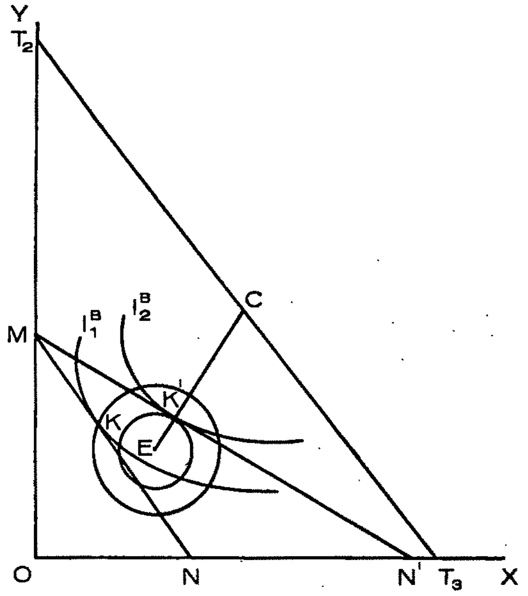


FIGURE 5

based on the assumption of only one consumption externality can lead us astray when we are, in fact, dealing with multiple consumption externalities.

It is clear from these examples that when we know only that the consumption of both X and Y by B confers external benefits to the rest of society, we cannot determine a priori whether a subsidy for X, a subsidy for Y, or a lump sum subsidy is required for efficiency. Indeed, no subsidy at all may be required since the competitive equilibrium will be efficient if (11) is satisfied. Only from a knowledge of the magnitudes of $MRS^A_{xz}/P_x/P_z$ and $MRS^A_{yz}/P_y/P_z$ can we determine the appropriate direction for policy to take. Knowing the mere existence of external benefits associated with certain goods is not enough. And, in the case of a large number of individuals who benefit from consumption of particular goods by B, there are well-known difficulties in obtaining information about their preferences since consumption by B exhibits the indivisibility of benefits and nonexclusion characteristics of a public good. What the analysis here emphasizes is the additional problem concerning the need for information about the preferences of externally affected parties relating *all* of the goods consumed by B before efficient policies can be designed.

An important qualification to these conclusions must be emphasized. Our analysis was based on the assumption that *all* of the goods consumed by B generate external benefits. When only some goods generate external benefits, the usual argument that competition leads to an underconsumption of these goods and an overconsumption of other goods is correct. Under these conditions a competitive equilibrium cannot be efficient, unless, of course, there are distortions other than these externalities present. But the information requirements for designing efficient subsidies are still tremendous.

Even if all goods consumed by B do not generate external benefits, the model developed here can still serve as a reasonable approximation if the proportion of B's budget devoted to goods with external benefits is high. To illustrate, let us compare, for a simple case, the quantitative difference between the external benefits generated by a lump sum subsidy (the policy recommended by our analysis as compatible with efficiency when goods generate external benefits to the same degree) and a consumption subsidy for particular goods (some form of which is the policy usually recommended to deal with external benefits in consumption). Assume that the consumption subsidy will increase the consumption of subsidized goods by the amount of the subsidy, and that the income elasticity of demand is unity for all goods. Suppose 10 percent of the consumer's budget is spent on goods with external benefits. A \$100 consumption subsidy will increase consumption of these goods by \$100 (assuming the appropriate goods are subsidized), but a lump sum subsidy will increase consumption by only \$10. The quantitative difference between the external benefits for the rest of society will then be quite significant, and the case for a consumption subsidy appears quite strong. But if 80 percent of the consumer's budget is expended on goods with external benefits, a \$100 lump sum subsidy will increase consumption of the target goods by \$80, fully 80 percent as much as the use of consumption subsidies. The quantitative ad-

vantage of consumption subsidies is much smaller, for as the percentage of the consumer's budget devoted to external benefit producing goods becomes larger, we approach the situation analyzed in this paper.[7]

Clearly, the argument for dealing with consumption externalities on a commodity by commodity basis rather than with a general cash subsidy is weaker the larger the proportion of the consumer's expenditures devoted to these goods. As noted earlier, this conclusion is particularly relevant to a discussion of policies designed to subsidize the poor since it has been often alleged that the consumption of a large number of important goods by the poor benefits the rest of society. If this allegation is true, then the theoretical case for subsidizing separately the consumption of these many goods is correspondingly weak.

[7] I do not mean to suggest that this is an adequate comparison of these policies since there are numerous other matters to take into consideration. For a fuller discussion, see the author.

REFERENCES

E. K. Browning, "Consumption Subsidies and Cash Transfers: The Controversy Revisited," unpublished paper 1972.

G. Daly and F. Giertz, (1972a) "Welfare Economics and Welfare Reform," *Amer. Econ. Rev.*, Mar. 1972, *62*, 131–38.

——— and ———, (1972b) "Benevolence, Malevolence and Economic Theory," *Publ. Choice*, Fall 1972, *13*, 1–19.

F. T. Dolbear, Jr., "On the Theory of Optimum Externality," *Amer. Econ. Rev.*, Mar. 1967, *57*, 90–103.

W. S. Vickrey, *Microstatics*, New York 1964.

Priorities and Pollution: Comment

By Abba P. Lerner*

In my article in this *Review*, "Priorities and Efficiency," I commented on the 1971 Report of the President's Council of Economic Advisors and discussed the relative advantages of limiting pollution by setting prices on pollution emission or by fixing the quantity permitted. There I argued that the line between the two methods is not very sharp. If the quantity is fixed, the price could still be set in a free market. If the certificates (or permits), each permitting a given amount of a particular pollution, could be freely traded, they would end up where the avoidance of the permitted pollution would have involved the greatest cost. The total cost of the given reduction in pollution would thus be minimized. The net result is the same as if this price had been fixed by the authorities instead of the quantity of permitted pollution.

I then indicated that a further improvement could be achieved by allowing conservationists to buy and sell different kinds of permits. They could then improve the nature of the reduction of pollution at negligible cost to themselves. I failed, however, to make clear that such an improvement in antipollution measures would still further cloud the distinction between the two methods—between fixing quantities and fixing prices. If such "trading" in pollution permits by conservationists is to be effective, the authorities cannot just issue fixed quantities of permits for each kind of pollution. They must provide more of one kind of permit in exchange for the cancellation of other permits of equal value at the market-determined prices.

For any given number of CO_2 permits and SO_2 permits, the market would set their prices at the levels which would equate the demand for each kind of permit with the supply. (The demand for the permits would come either from industries that emitted CO_2 or SO_2, or from conservationists who brought them up in order to reduce pollution.) Suppose the resulting market prices were \$10 for a permit for emitting a ton of CO_2 and \$20 for a permit for a ton of SO_2. If private conservationist groups believed that a ton of SO_2 did more than twice as much damage as a ton of CO_2, they could buy up SO_2 permits, trade them in at the government office for twice as many CO_2 permits, sell the CO_2 permits, and get all their money back. The authorities would have to cancel the SO_2 permits traded in and to print more CO_2 permits. Otherwise (i.e., if the authorities resold the permits they bought and rebought those they sold), this activity by the conservationists would have no effect. At this point it becomes difficult to say whether it is quantity or price which is being used in the antipollution program.

In the same issue of this *Review*, Jerome Stein criticizes the Report for suggesting that pollution controls would best be handled by local authorities, and insists that there be national regulation to equalize the charge for pollution damage throughout the nation. To the rather mystical argument that "the environment . . . belongs to the United States as a whole . . . " is added what looks like a more economic argument, viz., that " . . . optimal policy should require that the marginal product of the environment (in terms of producing real national income) be the same everywhere" (p. 535). He argues that if pollution damage per marginal unit of pollution is less in Phoenix than in Los Angeles, then if pollutive industries are transferred from Los Angeles to Phoenix, real national output (as conventionally measured) stays constant, while total pollution damage is decreased. Arizona loses less environment than California gains.

The argument for national, or at least wider, pollution regulation cannot be questioned in cases, also pointed out by Stein, where the localities would disregard damages

* Queens College, City University of New York.

to adjacent areas for which they are not responsible. But the main thrust of the argument seems to be directed at the possibility that "a unit of environment" might be valued differently by different local authorities. It is not clear, however, whether the loss of a unit of environment refers to some physical measure of pollution, or whether it refers to the valuations by the local population of the damage from the pollution that they suffer. In either case no wider control seems to be required.

If it takes ten tons of SO_2 to cause as much damage to people in Phoenix as one ton of SO_2 would cause to people in Los Angeles, because there are fewer people in Phoenix, then the charge for emitting ten tons of SO_2 in Phoenix should be equal to the charge for emitting one ton of SO_2 in Los Angeles. Wider control is unnecessary because the lower charge per ton emitted would tend to attract SO_2 emitting industries from Los Angeles to Phoenix.

But it may be that not only are there fewer people to be hurt by SO_2 in Phoenix, but the people there care less about the damage, so that they would, therefore, charge only one-twentieth as much as Los Angeles per ton of SO_2 emitted and there would seem to be different charges per "unit of environment used up" if we imagine that there is some objective measure of pollution damage other than how it is regarded by the people who suffer it. But in this case, too, there is no need for national intervention. There will be the appropriately greater temptation for SO_2 emitters to move from Los Angeles to Phoenix to take advantage of the greater difference in the pollution charge.

The movement of firms of this kind from Los Angeles to Phoenix will, of course, tend to reduce the difference in marginal pollution damage per ton of SO_2 and, therefore, in the prices of the permits. But there is no more reason for supposing that the difference must disappear than to suppose that the rent per acre of land must become the same in Phoenix as in Los Angeles because of people moving from Los Angeles to Phoenix. There appears to be a confusion here between a unit of pollution in some physical sense—a ton of SO_2—and a unit of *pollution damage* which ultimately can only be a subjective evaluation.

Stein does in several places explicitly say that it is the charge per unit of pollution *damage* that must be uniform nationally, and with this no quarrel is possible. But it is only the local population that feels the damage. A national regulation would have to have some objective measure of pollution *emission* to which the uniform charge would be applied. His expression unit of environment is uniquely fitted to obscure the difference between a unit of pollution and a unit of pollution damage.

While Stein's formula is unexceptionable in terms of pollution damage, he cannot avoid considering equalized "... effluent fees: charges for the emission of pollution ..." which would lead to "... an optimum distribution of a fixed quantity of pollution ..." (p. 532) only if every unit of pollution emission does the same amount of pollution damage. This might be the case locally, but not nationally. It is only by allowing the local authorities to impose *different* charges per (objective) unit of pollution emission that we could get some approximation to a *uniform* charge per unit of (ultimately subjective) pollution damage. The only national principle to be laid down (perhaps by the national authority) is that a dollar's worth of pollution damage must be charged a dollar (by the local authority).

It might seem that if some compromise is followed of fixing the total amount of some pollution damage to be permitted, then a uniform charge per unit of pollution damage would have to be imposed nationally. But there is no way of laying down any total amount of pollution *damage*, unless this is strictly proportional to some physical measure of pollution *emission*. If it was nevertheless decided to fix some total national amount of permitted pollution emission, the national authority could try to minimize the total national damage from the remaining pollution emission. Instead of imposing a uniform charge per unit of pollution emission (which

would minimize the cost of the pollution abatement), they would have to instruct each locality to charge the same percentage of the local marginal damage per unit of pollution emission. The national government would then have to keep adjusting this percentage to get it to bring about the previously decided upon total national amount of pollution emission—though it is hard to see why this would be of interest. It is easier to follow the optimal rule of instructing the localities to impose charges equal to 100 percent of the local pollution damage—a policy they would want to follow anyway.

REFERENCES

A. P. **Lerner**, "The 1971 Report of the President's Council of Economic Advisers: Priorities and Efficiency," *Amer. Econ. Rev.*, Sept. 1971, *61*, 527–30.

J. L. **Stein**, "The 1971 Report of the President's Council of Economic Advisers: Micro-Economic Aspects of Public Policy," *Amer. Econ. Rev.*, Sept. 1971, *61*, 531–37.

Priorities and Pollution: Reply

By JEROME L. STEIN*

The main difference between Abba Lerner and myself stems from our implicit social welfare functions. I regard the environment in a given locality as a public good from the point of view of all of the citizens of the United States, whereas he regards it as an argument appearing only in the utility functions of the residents of the given locality. This difference is crucial. On the other hand, the ambiguity concerning the distinction between "pollution" and "pollution damage" can be clarified without much difficulty. My conclusion is that optimization requires that there be a uniform national price for pollution damage. However there will not be uniform effluent fees among localities: prices per unit of pollution will differ among localities. This distinction was not sufficiently clear in my article. On the other hand, Lerner would leave the localities in complete charge of pollution control; and there would not be a single price for pollution damage among localities.

Problems in welfare economics are best understood by viewing them in the context of a democratic socialist society. A hypothetical socialist prime minister is concerned with environmental problems, and seeks advice from his economic advisors. What empirical information must he have in order to make a wise decision, and what value judgements enter into the final decision? By viewing the problem in this manner, the criticisms of Lerner and the other participants in the controversy can be placed in a proper perspective.

Suppose that there were two areas in a socialist economy, denoted by 1 and 2. Production in each area is described by equation (1a) where y_i is net output per capita after subtraction of rental payments rk_i, k_i is capital per capita, and X_i refers to the per capita nonlabor and noncapital inputs into the production process that affect the environment in this region.[1]

$$(1a) \qquad y_i = G^i(k_i, X_i) - rk_i; \qquad G_2^i > 0$$

The value of G_2^i is the loss of output that would result from a decrease in the rate of emission per unit of time, because resources with opportunity costs must be used to decrease the emissions resulting from the productive process.

It is desirable to derive a production function which relates net output y to input X exclusively. Firms in each region are assumed to employ a quantity of capital which will equate the value of the marginal product of capital $G_1^i(k_i, X_i)$ to the national rental price of capital r. This equality is described by equation (1b). Solve for k_i in terms of X_i and parameter r, and derive equation (1c).

$$(1b) \qquad G_1^i(k_i, X_i) = r$$

$$(1c) \qquad k_i = g^i(X_i, r)$$

Substitute (1c) into (1a) and derive equation (1d). The resulting production function is an envelope along which the marginal product of capital is equal to the national rental price r, and it relates net output y to the environmental input X exclusively.

$$(1d) \qquad y_i = G^i[g^i(X_i, r), X_i] - rg^i(X_i, r) = F_i(X_i)$$

Residents of each area realize that the use of X_i inputs into the production process is depriving them of something that is an argument in their utility functions. Production involves the use of input X_i; but this is not necessarily an argument in the typical resident's utility function. What is the argument in the utility function, other than y_i the

* Brown University and the Hebrew University, Jerusalem. I am indebted to Shlomo Yitzchaki for his perceptive comments on an earlier draft.

[1] My argument would not be affected if I worked with total, rather than with per capita, quantities.

quantity of the good consumed? It may be milligrams of sulfur dioxide per cubic meter of air, the loss of a view of nature or the dissolved oxygen content of the water. Define x_i as the argument in the typical household's utility function associated with the use of X_i in the production function. Preferences are defined on (y_i, x_i) whereas the supply set relevant for production is defined on (y_i, X_i). Variable x_i is what I mean by pollution damage, and X_i is what I mean by pollution or emission. Variable x_i is not measured in terms of utility but is the environmental argument in the utility function. The "loss of a unit of environment" is precisely what I meant by a unit of x_i. For example: The blaring of a loudspeaker may be an input X into the production process; but its effect upon the environment (x) depends upon whether the region is industrial or residential. Or, a given quantity of chemical deposited in the waterways (X) produces a different amount of damage if poured into a stream than it does in the ocean. The argument in the utility function is the purity of the water (x), not the quantity of chemical dumped (X). The distinction between products produced by firms and characteristics which enter the utility function, developed fruitfully by Kelvin Lancaster in his theory of demand, is extremely important in discussing environmental policies. My distinction between X and x is not a radical departure; but it is a means whereby meaningful theorems can be obtained.

Variable x_i is related to X_i in the manner described by equations (2a) and (2b).

$$x_i = H_i(X_i) \tag{2a}$$

$$X_i = H_i^{-1}(x_i) \tag{2b}$$

This equation is described in the lower part of Figure 1. If the residents of the ith region are to be able to decide upon the optimum

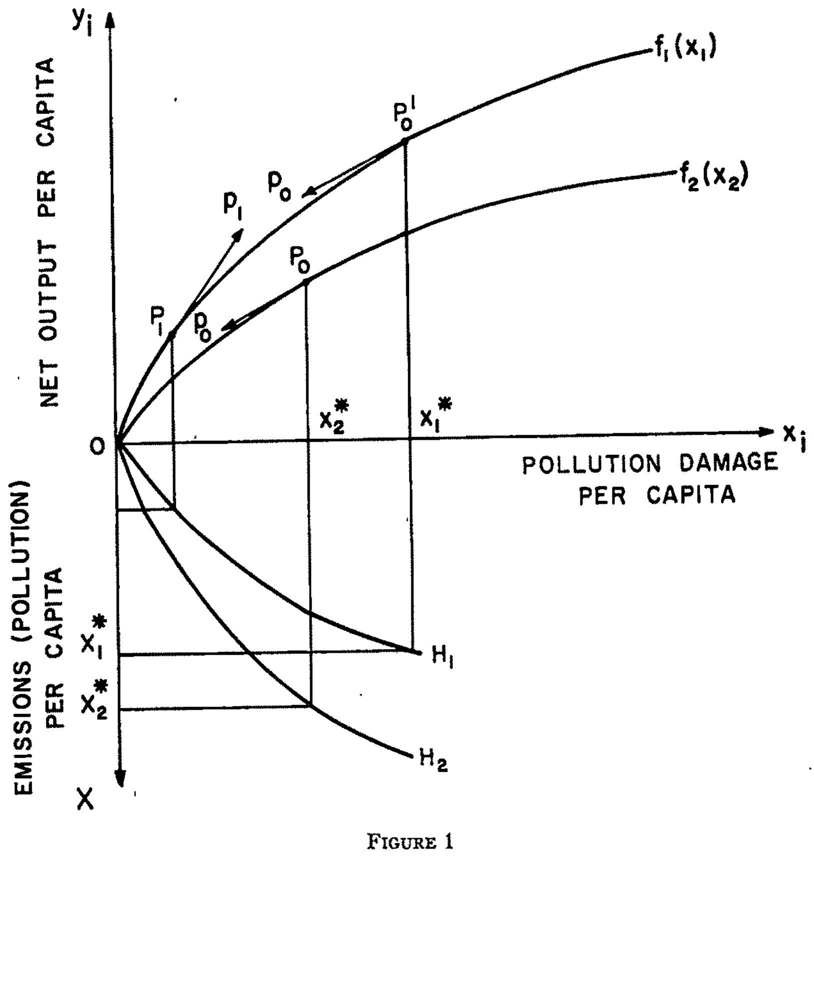


FIGURE 1

rate of pollution (X) in their own region, they must know the relation between the input into the production process (X) and its effect upon x, which is an argument in the utility function of the typical household. Function H, which maps X into x, will differ among localities, for the reasons cited earlier (for example, the relation between the dumping of a chemical and the purity of the water). Although the charge per unit of x may be uniform nationally, the charge per unit of pollution (emission) X will differ insofar as the H functions differ among localities. This was the step that was not clear to Lerner, Sam Peltzman and T. Nicolaus Tideman: an ambiguity noted by Allen Kneese. I apologize for not being sufficiently clear.

Substitute (2b) into (1d) and obtain a transformation function $f=(F \circ H^{-1})$ in each region, as described by equation (3).

$$y_i = (F \circ H^{-1})_i(x_i) = f_i(x_i) \tag{3}$$

which is drawn in the upper part of Figure 1. Now the production set is defined on the same elements as appear in the utility functions, and we can proceed to the optimization problem.

I. The Optimization Procedure Implicit in Lerner's Criticisms

Given the transformation function f, how should the prime minister decide what is the optimum amount of pollution? If Lerner is justified in claiming that "... it is only the local population that feels the damage," then the local utility function for the representative individual is equation (4).

$$V^i = V^i(y_i, x_i) = V^i[f_i(x_i), x_i] \tag{4}$$

Given the local transformation function f_i, if each locality maximizes its own utility function (4) separately, then the social welfare function implicit in Lerner's paper (described by my equation (5)) would be maximized.

$$W_L = nV^1[f_1(x_1), x_1] + (1-n)V^2[f_2(x_2), x_2] \tag{5}$$

when $1 \geqq n \geqq 0$ is a given weight. Presumably n is the fraction of the total population currently living in the first region. The optimal condition is that the marginal rate of substitution between output and environment $-\partial V^i/\partial x_i / \partial V^i/\partial y_i$ should be equal to the marginal rate of transformation $f_i'(x_i)$ in each region separately.

$$\frac{-\partial V^i/\partial x_i}{\partial V^i/\partial y_i} = f_i'(x_i) \tag{5a}$$

In region I, the indifference curve (not drawn here, but drawn in my original article) $V^1[f_1(x_1), x_1]=c$ would be tangent to $f_1(x_1)$ at point P_1; and in region II, the indifference curve $V^2[f_2(x_2), x_2]=c$ would be tangent to $f_2(x_2)$ at point P_0. Residents of region II are less concerned with the environment than are residents of region I. The price per unit of pollution damage, $p=f'(x)$, will differ among regions at the optimum point based upon Lerner's implicit welfare function (5).

I pointed out in my original article that localities may be prevented from optimizing, i.e., satisfying (5a), if they contain firms which have monopsony power. By threatening to relocate, the polluter could prevent the locality from moving to its preferred position. Hence local control may not even be optimal in terms of social welfare function (5).

Unfortunately, we are confronted with a deeper problem than is reflected by social welfare function (5); that is one reason why the entire nation is concerned with the issue of the environment. Suppose that: (i) the current residents of the coastlines decide to line them with huge billboards, junkyards, and garbage heaps; (ii) the upstream communities use the rivers and streams as waste receptacles; (iii) the current residents of California decide to cut down all of the redwoods. Are they the only ones that will feel the damage?

According to Lerner's implicit equation (4), the answer is "yes" in cases (i) and (iii). He did specify that, as in case (ii), local control is not optimal; and wider regulation would be desirable. My statement that "the environment belongs to the United States as a whole" was intended to mean that the

social welfare function is *not* of the form (5). The current ephemeral residents of the ith area are most definitely *not* the only ones concerned with its environment. Both residents of other areas and future generations have an interest in the environment of the ith region. That is, I consider the environment of the ith region as a public good insofar as the residents of *all* of the regions are concerned, whereas Lerner regards it as a public good *only* to the residents of the ith region. This is the crux of the difference between us. My answer to the questions (i)–(iii) posed above is "No." The utility function of the residents of the ith region would be equation (6).

$$V^i = V^i(y_i, x_i, x_j) \tag{6}$$

because people from one region travel through other regions for business and for pleasure. This equation reflects (among other things) the widespread objections to billboards, junkyards, and garbage heaps along public highways.[2]

II. The Optimization Procedure Implicit in My Public Good Case

The prime minister, representing all of the people, must have a social welfare function in the case where the environment in region i is a public good of concern to the residents of all of the regions. I selected social welfare function (7a) which has as its arguments two weighted sums: $y=ny_1+(1-n)y_2$ and $x=nx_1+(1-n)x_2$.

$$W = W(ny_1 + (1-n)y_2, \; nx_1 + (1-n)x_2); \qquad W_y > 0, \quad W_x < 0 \tag{7a}$$

My crucial assumption which reflects value judgements is that there exists an aggregative measure of environmental damage $x=nx_1+(1-n)x_2$. It is a weighted sum of the environmental arguments in utility functions (for example, water purity) in all of the regions. Then I can speak of maximizing total output $y=ny_1+(1-n)y_2$ associated with a given total amount of environmental damage; or of minimizing total environmental damage associated with a given rate of total output y.

Substitute transformation functions (3) into (7a) and derive social welfare function (7b).

$$W = W(nf_1(x_1) + (1-n)f_2(x_2), \; nx_1 + (1-n)x_2) \tag{7b}$$

The environmental problem confronting the prime minister is to select (x_1, x_2) which will lead to a maximum of social welfare. The solution is given by equation (8).

$$f_1'(\overset{*}{x_1}) = f_2'(\overset{*}{x_2})) = -W_x/W_y \tag{8}$$

The marginal rate of transformation between environment damage and output must be the same in each region.[3] This was the conclusion of my earlier article.

Figure 1 summarizes the normative argument. At the optimum (x_1^*, x_2^*) the marginal rate of transformation (denoted by p) between environmental damage x and output y must be the same in each area. The total quantities of output and environmental damage will depend upon the specific form of the social welfare function. Associated with $p_0=-W_x/W_y$ is a vector of pollution damage per capita in each region (x_1^*, x_2^*) and a vector of pollution or emissions (X_1^*, X_2^*).

The price per unit of pollution emission will not be the same in each region if the H functions differ between them. The price per unit of emission (pollution) will be described by (9).

$$p_0 \frac{x}{X} = p_0 \frac{H(X)}{X} \tag{9}$$

and it is higher in region I than it is in region II (as drawn in Figure 1).

What is the proper social utility function

[2] If there is an aspect of the environment in region i which only concerns the residents of this area, then the utility function would be (4) and the maximizing condition would be (5a) rather than (8). In that case, there would be no argument between Lerner and myself with respect to that aspect of the environment. Our disagreement arises because we are using different utility functions.

[3] The price of output is standardized at unity.

to be used? There are many possibilities and I have selected one which refers to the situation where residents of area i are concerned with the environment in area j, and which also yields clear-cut results. Lerner, on the other hand, assumes that the environment in area i is only of concern to the residents of that area. Obviously, the optimal condition will depend upon the type of social welfare function that is chosen; and this is the proper subject of controversy. This issue between Lerner and myself concerns the social welfare function, and not the distinction between pollution or emission X and pollution damage x.

III. Operational Principles

The principles which I have developed in equations (8) and (9) are no more difficult to implement than congestion tolls. The prime minister can implement equation (8) by selling a given quantity of freely transferable environmental usage certificates. Each certificate entitles the owner to produce one unit of x, pollution damage, anywhere in the country. As a result of arbitrage, the price per unit of damage will be uniform nationally. Call this market determined price p.

In consultation with the localities, the federal government determines a damage function $x_i = H_i(X_i)$ for each area. As described in the lower part of Figure 1, it is stated that a given quantity of emission produces more pollution damage in area 1 than it does in area 2. The H function tells us how many certificates are required per unit of emission in each area: it is similar to the relation between the number of vehicles on the road and a measure of congestion. In this manner, the total charge per emission px/X will be determined. Call this charge t_i, the product of p and x_i/X_i. For example, the H function may be:

$$x_i = a_i X_i^2, \qquad a_i > 0 \tag{10}$$

Then the price per unit of emission (pollution) is given by (10a) and will vary by region, insofar as the a_i's differ.

$$p\frac{x_i}{X_i} = p a_i X_i \equiv t_i \tag{10a}$$

Price p is determined in the national market. Ratio x_i/X_i reflects the country's evaluation of local conditions and is similar to a measure of local congestion. As the quantity of emissions rises, more environmental usage certificates are required per unit of emission since $x_i/X_i = a_i X_i$. The national market establishes p (the price per unit of "congestion") and the locality announces the current x_i/X_i (the level of congestion per automobile), based upon the H function agreed upon by the federal government and the locality. The rising supply price t_i of emissions X_i will differ among localities insofar as the a_i's and X_i's differ. Both p and x/X are variables.

Firms will try to maximize profits π and will use that quantity of emissions which equates the value of the marginal product of pollution $F_i'(X_i) = G_2^i$ to the price t_i per unit of emission. This is described by equations (11a) and (11b), based upon equations (1a)–(1d). At all points on F_i, the value of the marginal product of capital is equal to the national rental price of capital.

$$\pi = G^i(k_i, X_i) - rk_i - t_i X_i \tag{11a}$$

$$\frac{d\pi}{dX_i} = (G_1^i - r)\frac{dk_i}{dX_i} + (G_2^i - t_i) = 0 \tag{11b}$$

Since $(G_1^i - r) = 0$ from equation (1b), it follows that:

$$G_2^i = F_i'(X_i) = t_i \tag{11c}$$

Insofar as the H functions differ among localities, the prices t_i per unit of emission (pollution) X_i will differ. If emissions produce a "lot" of damage in area 1 but "less" damage in area 2 (i.e., $H_1(\overline{X}) > H_2(\overline{X})$), then t_1 will exceed t_2. Firms in each area will adjust the quantities of capital per capita k_i and environmental inputs per capita X_i so as to satisfy two relations: (a) the value of the marginal product of capital will equal the national rental price r; (b) the value of the marginal product of pollution $F_i'(X_i) = G_2^i$ will equal the local price per unit of emission t_i. The higher the local price t_i per unit of emission, the lower will be the rates of output per firm and the number of firms operat-

ing in the locality. The emissions purchased per unit of x will be a decreasing function of the total emissions in the area, according to the H function (10). This is equivalent to varying the highway toll as a function of congestion. I do not see any serious problem in implementing this scheme.

There will be industrial or pollution concentration where a_i is relatively low, since input price $t_i=(a_iX_i)\ p$ will be relatively low. Where emission produces a considerable amount of environmental harm, a_i and hence t_i will be relatively high. Firms would not find it profitable to locate in those areas. Population will distribute themselves according to their (y, x) preferences. According to Figure 1, area 2 will contain people who prefer environment and area 1 will contain people who prefer goods, relative to the other commodity.

By varying the total quantity of environmental usage certificates, $\sum n_i x_i = x$, the prime minister can affect the total amount of pollution damage in the country. The total will be distributed among the areas in the optimal way, as described by equation (8). Then, the prime minister can decide if the total x in the country is too large or too small. This is a trial and error way of deciding upon the optimum total amount, when functions f_i are not known by the government. Given the amount of environmental damage x, the total output y will be maximized.

This is a rigorous presentation of what I attempted to do in my original article. I hope that the implicit assumptions, ambiguities, and differences between Lerner and myself have been clarified.

REFERENCES

A. V. Kneese, "Pollution and Pricing," *Amer. Econ. Rev.*, Dec. 1972, *62*, 958.

K. Lancaster, "A New Approach to Consumer Theory," *J. Polit. Econ.*, Apr. 1966, *74*, 132–57.

S. Peltzman and T. N. Tideman, "Local versus National Pollution Control: Note," *Amer. Econ. Rev.*, Dec. 1972, *62*, 959–63.

An Analysis of Turning Point Forecasts: A Fairly Polite Comment

By Rendigs Fels and C. Elton Hinshaw*

Once upon a time, a researcher was expected to look up previous work on his subject. In this more civilized age, he is spared the trouble, and if he inadvertently repeats the investigations of his predecessors, he gains credit for independent discovery. The predecessors gain too, for they get an opportunity to call attention to their forgotten findings. We are accordingly grateful to H. O. Stekler for overlooking our little book in his article in this *Review* (and even more grateful to the anonymous referee whom Stekler thanked for "valuable comments," p. 724). Geoffrey H. Moore and Julius Shiskin have reason to be grateful, too. Stekler could hardly have been unaware of their voluminous writings on the economic indicators and turning point forecast records, but he managed to discuss the subject without mentioning their works and excluded them from his list of references.[1]

Since previous investigations went well beyond Stekler's, we shall not attempt to recapitulate them. But parts of Stekler's note require comment, particularly his leading finding: "On the basis of this evidence, we may conclude that the failure to predict or identify cyclical peaks may be attributable to the forecasters extremely low subjective probabilities of the likelihood of such an event" (p. 729). His meaning is less vacuous than the quotation sounds. He assumed that "... forecasters begin with some subjective probabilities about the likelihood of a turn. As new information becomes available these subjective probabilities are then revised" (p. 724). He illustrated by assuming that a forecaster began the year 1957 believing the chances of a cyclical peak were 10 percent. If the forecaster had looked only at the Federal Reserve Board (*FRB*) Index of Industrial Production, he would have revised the estimate upward to 25 percent in February, 50 percent in March, and 74 percent in April. A peak actually occurred in July. A similar calculation for 1960 gave similar results but with a shorter lead. From these and other exercises, Stekler inferred that forecasters' initial subjective probabilities of a cyclical peak were "extremely low" (p. 729). Extremely low apparently means well below 10 percent, perhaps as low as zero (see p. 727).

Given the number of indicators and the extensive analyses of their historical behavior, the choice of the Index of Industrial Production, a roughly coincident series, seems odd. It led at the peaks of 1957 and 1960 but not at those of 1953 and 1969. Of the five other series Stekler studied (p. 728), two are leading indicators, three roughly coincident. There are seventy-two economic indicators of the National Bureau of Economic Research (*NBER*), fifty-one of which have been classified into leading, coincident, or lagging groups. In the *NBER's* short list, there are twenty-six indicators—twelve leading, nine roughly coincident, and five lagging. The smallness of the sample used by Stekler both in terms of number of series and number of peaks is open to question. In these respects, Stekler's work suffers in comparison to that of Moore and Shiskin.

Shiskin, for example, reviewed the be-

* Professor of economics, and associate professor of economics and associate dean, College of Arts and Science, respectively, Vanderbilt University. We gratefully acknowledge the helpful comments of Julius Shiskin.

[1] Although Stekler stated that "There have been a number of empirical analyses of the accuracy of economic forecasts . . . ," he cited only two studies (Henri Theil and Victor Zarnowitz) in addition to his own in support of this statement, p. 724. His list of references also included Michael Evans et al. The omission of Moore's studies is especially remarkable, since his Presidential Address to the American Statistical Association in 1968 reviewed and appraised the performance of forecasters in predicting and recognizing turning points and referred to Stekler's papers on the subject.

havior of the indicator indexes (lagging inverted, leading, coincident, lagging, and diffusion) around cyclical turns from 1920 to 1960 (pp. 89–114). In contrast to Stekler, who reviewed only five indicators and found that if forecasters had prior probabilities of greater than zero, they would have identified the two turns quickly (p. 727), Shiskin argued that the signs of a business cycle peak in July 1957 were no more decisive than in mid-1965 (pp. 108–09), and that the May 1960 peak was difficult to recognize because of the "relatively slight changes in the level of activity during the whole period from February to July" (p. 111). Shiskin emphasized that the indicators must be used with other data and that one must be aware of their limitations (p. 114).

Whereas Stekler inferred forecasters' subjective probabilities indirectly, we studied them directly. We studied monthly or weekly forecasts published in eight periodicals for the eight cyclical turns (four peaks and four troughs) between 1948 and 1961, forecasts in the minutes of the Federal Open Market Committee (*FOMC*) for the entire period 1947–60, and forecasts of two publications relying heavily on the *NBER* leading indicators for the two peaks and two troughs between 1957 and 1961. We used two scoring systems, one representing the subjective probability of the forecaster concerning the imminence of a business cycle turning point ("certainty score"), the other representing the accuracy with which he pinpointed the *NBER* date. We also studied false alarms. For 1957 our evidence contradicts Stekler's expectation. We found that "warnings of recession were less frequent three months before the peak than they had been earlier . . . " (p. 30). Instead of the forecasters starting with extremely low subjective probabilities and revising them upward, they started with unusually high expectations and revised them downward. This was true until April 1957, the terminal month of Stekler's exercise. The certainty scores increased slowly during the next several months and spurted sharply in November (see the authors, p. 30). In general our certainty scores (equivalent to Stekler's subjective probabilities) are too high to be consistent with his conclusion. This is also true of 1960, though to a lesser degree than 1957. The scores are, however, low enough to confirm Stekler's observation that forecasters are generally unsuccessful in predicting cyclical peaks.

Why do forecasters have a poor record of forecasting cyclical turns? We reject Stekler's answer because it is disconfirmed by our evidence. Since Stekler's empirical work was confined to 1957 and 1960,[2] we shall confine our comments to them also. Aside from having been burned in 1956 when a number of them predicted a recession that did not occur (see the authors, pp. 37–38), forecasters in 1957:

> were misinformed about investment in inventories, investment in plant and equipment, and federal spending.
>
> . . . .
>
> Forecasters always have to work with inaccurate information. Usually some of the errors offset each other. In this case, they all worked in the same direction. Of the three errors, one was of decisive importance. In May the Department of Commerce had published an estimate for the first quarter showing a substantial negative figure for inventory investment, leading forecasters to think the economy had already weathered an inventory adjustment. Revised figures now show substantial positive inventory investment. [p. 31]

Concerning 1960 we said:

> A bulge in inventory investment was expected to give the economy a strong upward thrust in the first half of 1960. The aftermath of the restocking period was expected to be a decline in inventory investment with unfavorable repercussions on the economy as a whole.

[2] In a footnote Stekler said, "There has as yet been no general analysis of the 1968–70 *GNP* forecast errors. It would be interesting to determine whether this peak was again not predicted or 'identified' " (p. 729). Mary Virginia Hill has applied our scoring method to a similar group of forecasts made in the vicinity of the 1969 peak. She found that " . . . forecasters performed relatively well at the 1969 peak, scoring second highest for accuracy and certainty among post-World War II peaks" (p. 44). In Stekler's terms, "relatively well" means not very good.

> But forecasters drew the conclusion that the outcome would be a slower rate of expansion rather than an immediate downturn, with strong possibility of a recession beginning late in 1960 or early in 1961. Their logic seems at fault. Their diagnoses implied that the danger point would come at midyear rather than at year's end.[3] [p. 33]

For a remedy, Stekler suggested that forecasters "pay more attention to the statistical indicators, especially the leading series" (p. 729). We concur, but a qualification is in order. Our findings verify that those forecasters who relied heavily on the indicators had a better record of forecasting, or at least identifying, the peaks of 1957 and 1960 than the others, except for the *FOMC* (see the authors, pp. 41, 43, 44, 93, and 100).[4] We also found that they did worse at the troughs of 1958 and 1961 (pp. 47–48), but no matter. The chief drawback to Stekler's suggestion is false alarms. In a sense, Stekler took account of false alarms in his procedure for calculating the rise in subjective odds from an initial 10 percent. "In the period 1947–56 there were 12 movements when the Index [of Industrial Production] was below its previous peak for one or more months but which were not followed by recessions. These movements occurred in 31 of the 95 months in which there was no recession" (p. 725). This experience was the basis for calculating the rise in subjective odds from observing the Index during the first months of 1957. So far, so good. But Stekler did not follow through. He should have applied his methods to episodes when no recession occurred. It is true that the leading indicators have given early warning of every business cycle peak, but they have also given early warning of some that never occurred.[5]

If the definition of a business cycle recession is modified to include a slowing down or leveling off in the rate of growth, our criticism about false alarms loses its sting. Declines in the leading indicators index have always signaled a retardation in growth, some of which developed into recessions in the traditional sense. But as Moore pointed out, "Recognizing business recessions is not the same as recognizing every wiggle on the curve. Major lasting movements must be distinguished from minor interruptions of the trend" (1969, p. 9).

Stekler would have done well to heed Shiskin's warning that the behavior of the indicators ". . . does not supply an automatic or explicit forecast of changing business conditions, nor will it pinpoint the month of a cyclical turning point in advance. It can be a valuable statistical tool in the hands of a skillful and well-informed analyst of changing business conditions. It is not a mechanical guide to the future" (p. 15).

[3] A footnote adds:

> The forecast of continued expansion throughout 1960 was not unreasonable. The upper turning point was so flat and the ensuing contraction was so mild that, one may infer, absence of any of the deflationary forces actually at work would have prevented the cyclical turn. One of the factors at work was the failure of consumer spending to rise in the third quarter in spite of a rise in disposable personal income. This development and its consequences could hardly have been predicted with the forecasting tools available in 1960 (or now, for that matter). Nevertheless, the forecasters ought to have been able to give clearer warning to the effect that, though the outlook was favorable, the odds in favor of recession were by no means negligible. [p. 33]

[4] Hill likewise concluded that at the 1969 peak, "The analysts who relied heavily on statistical indicators proved more sensitive than the average eclectic forecaster" (p. 44).

[5] In particular, the indicators misled forecasters in 1956 and 1962–63 (see the authors, p. 46). "One forecaster relying primarily on business cycle indicators in May 1956 was rather sure a cyclical peak had occurred in February" (p. 37). And Shiskin found that "most of the leading business indicators pointed to a reversal of the upward trend of the economy in 1951" (p. 108).

REFERENCES

M. Evans, Y. Haitovsky, and G. I. Treyz, "An Analysis of the Forecasting Properties of U.S. Econometric Models," in B. Hickman, ed., *Econometric Models of Cyclical Behavior*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 36, New York 1972.

R. Fels and C. E. Hinshaw, *Forecasting and Recognizing Business Cycle Turning Points*, New York 1968.

M. V. Hill, "Evaluation of Forecasts of

Turning Points in the Business Cycle: the 1969 Peak," M.A. thesis, Vanderbilt Univ. 1972.

G. H. Moore, *Business Cycle Indicators*, Princeton 1961.

———, "Forecasting Short-Term Economic Change," *J. Amer. Statist. Assn.*, Mar. 1969, *64*, 1–22.

——— and J. Shiskin, *Indicators of Business Expansions and Contractions*, New York 1967.

J. Shiskin, *Signals of Recession and Recovery, An Experiment with Monthly Reporting*, New York 1961.

H. O. Stekler, "An Analysis of Turning Point Forecasts," *Amer. Econ. Rev.*, Sept. 1972, *62*, 724–29.

H. Theil, *Applied Economic Forecasting*, Chicago 1966.

V. Zarnowitz, *An Appraisal of Short-Term Economic Forecasts*, Occas. Paper 104, Nat. Bur. Econ. Res., New York 1967.

An Analysis of Turning Point Forecasts: A Polite Reply

By H. O. STEKLER*

I apologize to Rendigs Fels and C. Elton Hinshaw for having failed to footnote their book. At the same time I am sure Geoffrey Moore and Julius Shiskin will forgive me for not having noted their work on the leading series for I came not to bury the leading series but rather to praise them. This was in sharp contrast to my earlier writings (which I also didn't reference) in which I was at least somewhat critical of this approach, because of the large number of false leads.

Fels and Hinshaw claim that their work measures forecasters' subjective probabilities directly, and this assertion should be examined. Although this is not the appropriate time to review a book I first read five years ago, some comments are in order.

They cannot directly measure the subjective probabilities, for they only look at the words of the published predictions and from that language attempt to infer what the forecaster had in mind. Given their interpretations, they then assign certainty scores to these forecasts. I doubt that these scores should be called subjective probabilities, for we have no way of knowing whether these scores accurately reflect the state of the forecaster's mind; Fels and Hinshaw themselves indicate that their scores should not be interpreted too precisely, p. 21. Moreover Fels and Hinshaw indicated that their scores differed not only between themselves, but also each one gave different scores when he looked at the same forecasts at different instances of time, p. 20. These are limitations in their analysis.

Nevertheless, their results do show an increase in certainty (as they measure it) with time, but at a much slower rate than would have been predicted by the type of analysis I presented. It would be interesting to conjecture why this occurred, but they never directly compare the movements in the certainty scores with changes in the objective contemporaneous data. Thus they do not determine why forecasters did (did not) change their minds in any given month. They conclude that forecasters generally don't recognize the peaks but do not present an overall hypothesis to explain this phenomenon.

On the other hand, the analysis I have presented develops a set of posterior probabilities based upon objective data and alternative states of prior (subjective) probabilities. It is a process which may be applied to any set of objective data and the results would be the same regardless of who did the investigation. In other words, the technique is reproducible, hence, scientific. Consequently, it is possible to set up a number of different experiments and attempt to determine why turning point errors occur.

In order to improve the state of the art, it is not sufficient to report that turning point errors occurred. It is also necessary to determine why they happened. Only then would it be possible to correct predictive procedures and improve the quality of the forecasts.

For instance, in my earlier paper, I conjectured that forecasters had asymmetrical priors regarding downturns in *GNP* on one hand and decreases in the rate of increase in price indexes on the other hand. Using this Bayesian approach, Marjorie Schnader tested this hypothesis. She used signals in the Consumers Price Index (*CPI*) itself to predict further declines in that index, just as signals in the Federal Reserve Board (*FRB*) Index were used in the earlier analysis. Her results (Table 1) showed that the *CPI* probabilities were similar to the *FRB* probabilities. However in 1969 and 1970, forecasters repeatedly forecast that inflation would end shortly. These were strong predictions. Given their objective statements and the re-

* State University of New York, Stony Brook.

TABLE 1—PROBABILITIES OF SIGNALS AND NO SIGNALS IN THE *CPI* FOR TURNS AND NO TURNS IN THE *CPI*, GIVEN PRIOR PROBABILITIES OF TURN AND NO TURN, 1948–68

State of *CPI*	Information from *CPI* Signal	No Signal
Turn	1.00 (p)	0.0 (p)
No Turn	0.41 ($1-p$)	0.59 ($1-p$)

TABLE 2—POSTERIOR PROBABILITIES, GIVEN A SIGNAL IN *CPI*, FOR VARIOUS PRIOR PROBABILITIES

$P(T)$	$P(T/S)$
0.0	0.0
0.10	0.21
0.20	0.38
0.30	0.51
0.40	0.62
0.50	0.70
0.60	0.79
0.70	0.85
0.80	0.91
0.90	0.96
1.00	1.00

sults of Table 1, it is possible to infer what the prior probabilities must have been to obtain such high posterior probabilities. Table 2 shows that p must have been equal to or greater than 0.50 to get posteriors higher than 0.70. This is in sharp contrast to the priors which were close to zero for predictions of downturns, and confirms the presence of asymmetries in forecasters' priors. Consequently, this Bayesian approach enabled us to test a hypothesis about forecasters' expectations.

Turning now to Fels and Hinshaw's minor points. The *FRB* Index, even though it was coincident in 1953, was chosen to determine why forecasters fail to predict or to *recognize* cyclical peaks. To recognize a coincident cyclical turn, one need not use a leading series. Only a subset of the leading series were examined in conjunction with the *FRB* Index because the approach was intended to be illustrative and I merely wanted to show that *a representative* set of other data was not inconsistent with these movements in the *FRB* Index. Finally, I did not examine false movements because I was only interested in determining why true peaks were not forecast or recognized.

REFERENCES

R. Fels and C. E. Hinshaw, *Forecasting and Recognizing Business Cycle Turning Points*, New York 1968.

M. Schnader, "Forecasting Peaks in the CPI Index," working paper series no. 116, Stony Brook.

H. O. Stekler, "An Analysis of Turning Point Forecasts," *Amer. Econ. Rev.*, Sept. 1972, *62*, 724–29.

Transport Costs and the Static Welfare Costs of Tariffs

By W. G. Waters II*

An object of tariff policy is to raise the domestic price of foreign goods to give an advantage to domestic suppliers. An f.o.b. tariff applies to the price in the foreign country i, p_i, while a c.i.f. tariff is applied inclusive of international transport costs f_i. The domestic selling price will also include any domestic transport costs F_i and any nontransport costs associated with movement NTC_i (for example, costs of holding inventories because of the delay in replenishing stocks from distant suppliers). Given alternative suppliers, routes, and transport modes with varying p, f, F, and NTC, the imposition of tariffs can cause imports after protection to come from other than the true lowest cost supplier.[1] If this substitution occurs there is a deadweight welfare loss in addition to the usual production and consumption effects (henceforth P and C effects, respectively). This we might call the transportation effect (henceforth T effect).

I. The Transportation Effect

The easiest illustration of the T effect is a situation where only f.o.b. valuation can cause a distortion. Consider a small country which faces perfectly elastic suppliers at delivered price p_i+f_i (let us ignore F_i and NTC_i temporarily).[2] Freight costs are assumed constant per unit of commodity. In Figure 1, the home country initially (free trade) imports M from the lower delivered cost supplier, country 1. Assume $p_1>p_2$, $f_1<f_2$, and $(p_1+f_1)<(p_2+f_2)$. The country imposes an f.o.b. tariff and the domestic price rises (though by less than the amount of the tariff due to the geographic substitution). The usual static welfare loss due to a tariff is measured by the shaded triangles ABC and DEF. If this geographic substitution in favor of more distant suppliers does arise, there is the additional deadweight loss (the T effect) of $CC'D'D$ or $\{(p_2+f_2)-(p_1+f_1)\}M'$ where M' is the level of protected imports. For the same protection the only difference between the c.i.f. and f.o.b. valuation is the additional deadweight loss entailed by the f.o.b. base. Part of what would have accrued to the home country government as customs revenue under a c.i.f. tariff is instead paid for increased transportation costs under the f.o.b. method (actually the outlay for freight costs is higher than this net loss but it is partially reduced by purchasing imports from a lower foreign cost producer). The Appendix contains a calculation of the maximum size of the T effect relative to the P and C effects for an "average" commodity; the T effect appears to be quite small.

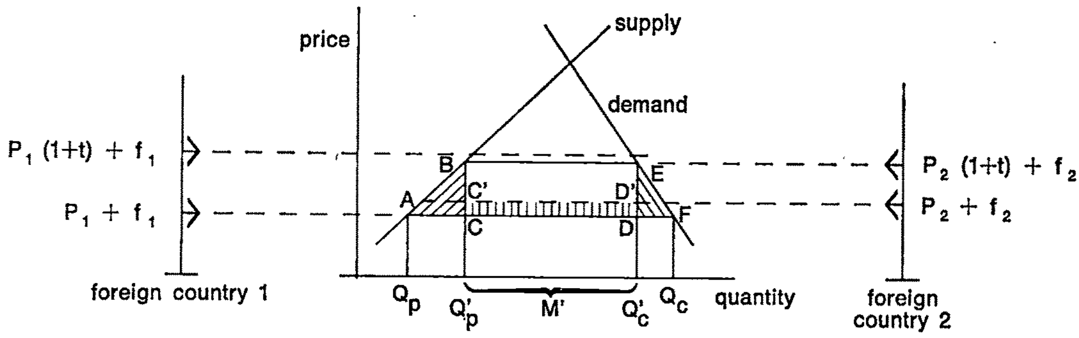
Suppose, instead, that Figure 1 refers to a single inland site served by alternative inland transport routes. Further assume ocean freight costs are identical and included in p_1 and p_2, then f_1 and f_2 would represent inland freight costs. The f.o.b. distortion is the same as before only the loss is associated with increased use of domestic haulage. Under these circumstances, the use of a c.i.f. base gives rise to almost the same T effect.

It is possible for the c.i.f. valuation to result in a T effect where the f.o.b. would not. The simplest example is where a "small"

* University of British Columbia. I would like to thank Jagdish Bhagwati, an anonymous referee, the managing editor, and my colleagues W. E. Diewert, Michael Kelly, Philip Neher, and A. D. Woodland for their helpful comments and suggestions.

[1] Substitution can be between different suppliers with unequal international freight costs (see Harry Johnson). It is sometimes possible to substitute domestic for international transport by choice of different routes (see my 1971 paper). Faster and higher quality transport modes involve substituting higher freight rates for lower NTC (see R. Maurer, Peter Diamond and Frank Mitchell).

[2] This is simply a convenience, the same results are obtained if we assume F_i and NTC_i do not vary among suppliers or routes.



Deadweight Loss of an F.O.B. Tariff

Production Effect ABC
Consumption Effect DEF
Transportation Effect CC'D'D

FIGURE 1

country imports from a single, large foreign supplier directly by ocean transport f_1 or via a shorter ocean route f_2 plus some inland freight costs F_2. Initially $f_1<(f_2+F_2)$ therefore goods are shipped by the more lengthy ocean route. An f.o.b. tariff imposed on the same foreign price would not alter the relative attractiveness of the two routes but a c.i.f. tariff t_c might. It is possible for $(p+f_1)\cdot(1+t_c)>\{(p+f_2)(1+t_c)+F_2\}$. Figure 1 could be adapted to illustrate this.

The valuation method can also affect the choice of international transport mode. Protecting domestic producers by raising the price of imports leads to a (partial) rise in *NTC* (for example, inventory and insurance costs rise with the increased value of the product). High-cost, high-quality modes offer increased f for decreased *NTC*, thus f.o.b. valuation encourages the use of such a mode. The c.i.f. valuation reverses this bias because the tariff is levied on the embodied international freight charges thus inflating them relative to the rise in *NTC*.

II. Conclusion and Implications

The above are particular illustrations. More generally, the possibilities of substituting among different suppliers, transport modes, and international and domestic routes are simultaneously present. The f.o.b. method encourages substitution to a more-distant, lower-price supplier. The increased transport expenditures may be international or domestic, with some bias for high-cost, high-quality international modes. The c.i.f. method encourages substitution of domestic for international transport and the latter is biased in favor of lower-cost, lower-quality service.

The appropriate valuation base appears to be an empirical question, although I would guess that c.i.f. valuation is preferred for most countries. Both valuation methods can distort the choice of international transport mode and there is, therefore, no a priori preference here. But most countries probably have limited opportunity to substitute domestic haulage for international (the c.i.f. bias) whereas their trading partners lie at various distances away (therefore the f.o.b. distortion of Figure 1 is a likely occurrence).[3]

There is one general rule for tariff policy. The *T* effect arises if protected imports come from other than the true lowest cost supplier. For the same (*ex post*) increase in the domestic price, tariff revenues will be less by an amount equal to the deadweight loss of the *T* effect. Therefore the appropriate method

[3] One is reluctant to accept Robert Baldwin's advice that countries should be adopting the f.o.b. method of Australia, Canada, and the United States, pp. 138–39, 183–84.

TABLE 1—MAXIMUM DEADWEIGHT LOSS OF THE T EFFECT RELATIVE TO THE P AND C EFFECTS
(using assumed representative figures)

Domestic Price Elasticities[a]	$\frac{Q'_P}{M'}=2,\ \frac{Q'_C}{M'}=3$	$\frac{Q'_P}{M'}=3,\ \frac{Q'_C}{M'}=4$
$e_s=1,\ e_d=-1$	3.0%	2.2%
$e_s=.5,\ e_d=-.3$	8.8%	6.3%
$e_s=.5,\ e_d=-.5$	6.4%	4.6%
$e_s=.25,\ e_d=-.5$	7.6%	5.6%

[a] Elasticities of unity were arbitrarily chosen. Those in the second row are John Floyd's aggregate elasticities for the United States. The third and fourth row are elasticities for "durable finished manufactures" and "nondurable finished manufactures" for the United States (Robert Stern).

is the one which yields maximum tariff revenue consistent with achieving, *ex post*, a specified increase in the domestic price of importables.

Use of specific tariffs or quotas will avoid most examples of the T effect except for a bias in favor of high-cost, high-quality modes because NTC rises with the increased price of the protected product. Only a direct subsidy can prevent the T effect altogether.[4]

APPENDIX

A general impression of the relative magnitudes of the T effect can be obtained by manipulating the formulae for the welfare costs of the P, C, and T effects and inserting some reasonable values for the variables involved. Return to Figure 1 (the case of f.o.b. distortion where F_i and NTC_i were omitted). The welfare costs of the T effect W_T relative to the welfare costs of the P and C effects, W_P and W_C, is given by

$$W_T/(W_P+W_C)=\{(p_2+f_2)-(p_1+f_1)\}\div \left(t^2(p_1+f_1)\left\{\frac{e_s}{1+te_s}\frac{Q'_P}{M'}-\frac{e_d}{1+te_d}\frac{Q'_C}{M'}\right\}\right)$$

where Q'_P/M' is the ratio of domestic production relative to imports under protection, Q'_C/M' is domestic consumption relative to imports, e_s and e_d are the elasticities of domestic supply and demand, and the other terms are already defined. This expression is in a useful form because there is no need to measure the absolute volume of trade and there are estimates for most of the remaining variables. For example, Q'_C/M' lies between 2 and 4 for most countries, 3 and 4 are used for purposes of illustration; this entails a value of 2 and 3 for Q'_P/M'. Arbitrarily assign a value of 20 percent for the tariff and a value of unity for price p_1. The average size of freight costs relative to product value is on the order of 10 percent, therefore we employ figures close to this, specifically .08 and .12 for f_1 and f_2, respectively. Given figures for p_1, f_1, f_2, and t, this determines the maximum value p_2 can assume and still give rise to a T effect.[5] Table 1 lists the maximum value the T effect can have relative to the P and C effects for various elasticities of supply and demand. The size of the T effect is quite small for the average commodity. It is sensitive to the difference in freight costs; if they are doubled (for example, $f_2=.16$), the sample calculations are doubled. Adding F_i+NTC_i

[4] There might be some increased transport costs on domestically produced goods. But this is part of the higher costs of domestic producers and is different from the distortion among foreign suppliers which is the T effect.

[5] The larger is p_2 the lower is the f.o.b. tariff necessary to yield the desired domestic price increase of 20 percent. The lowest possible f.o.b. tariff is the one which would be imposed on p_1 to yield the same protection as a c.i.f. tariff. In this case, $p_{2_{\max}}=.9671$ and the maximum per unit (of import) deadweight loss is .0071, i.e., less than 1 percent of the costs of production of actual imports.

to p_i+f_i alters the order of magnitude in Table 1 only if the difference in total delivery costs between the two valuation methods is greater or less than the difference in ocean freight costs postulated in the above example (.04 relative to a delivered price before protection of 1.08). Since the above figures represent some "average" commodity, the maximum size of the T effect could be less (zero) but could be several times larger for particular commodities with alternative suppliers, routes, and modes.

REFERENCES

R. E. Baldwin, *Nontariff Distortions of International Trade*, Washington 1970.

P. A. Diamond and F. Mitchell, "Customs Valuation and Transport Choice," *J. Int. Econ.*, Feb. 1971, *1*, 119–26.

J. E. Floyd, "The Overvaluation of the Dollar: A Note on the International Price Mechanism," *Amer. Econ. Rev.*, Mar. 1965, *55*, 95–107.

H. G. Johnson, "A Note on Tariff Valuation Bases, Economic Efficiency and the Effects of Preferences," *J. Polit. Econ.*, Aug. 1966, *74*, 401–02.

R. Maurer, "Customs Value and Inclusion of Transport Costs in That Value," Study no. 1969/10-E, Institut du Transport Aerien, Paris 1969.

R. M. Stern, "The U.S. Tariff and the Efficiency of the U.S. Economy," *Amer. Econ. Rev. Proc.*, May 1964, *54*, 459–70.

W. G. Waters II, "A Note on Transport Costs and the Choice of a Tariff Valuation Base," *J. Polit. Econ.*, July/Aug. 1971, *79*, 926–28.

The Theory of the Firm in "Long-Run" Equilibrium

By EUGENE SILBERBERG*

In a recent series of articles,[1] economists have investigated the comparative statics of the "neoclassical" firm in a competitive environment with the specific proviso that output price adjusts in response to changes in factor prices. The older theory of the firm, such as is found, for example, in Paul Samuelson (1947), is deficient in this regard. Metaphors such as "short run" have been used to gloss over the inconsistency of holding output price constant in the face of changing costs through changing factor prices. In this paper, a simple yet thorough analysis of a firm in a competitive industry, i.e., a firm facing zero entry and adjustment cost conditions, is given. New results pertaining to restrictions on the behavior of such firms are offered. Also, the length and complexity of established results is substantially reduced.

I. The Model

Consider a single firm whose productive process is summarized by the production function $y=f(x_1, \ldots, x_n)$, where y is output and $X=(x_1, \ldots, x_n)$ is a vector of inputs. The firm is a "price taker" in all markets, purchasing its inputs x_i at constant unit factor prices p_i, respectively, and selling its output at constant unit price p. However, p is not to be determined exogenously, as is the case in, say, Samuelson (1947). With zero entry and adjustment costs,[2] output price will always be bid down to the minimum average cost (min AC) level of the marginal firm in the industry. If we assume that all firms are identical in the sense of having identical min AC levels, i.e., that the industry supply curve is horizontal,[3] then all firms in the industry will produce at the point of minimum long-run average cost. That is, the profit maximizing policy of a firm under these conditions is to choose that mix of factors at which AC is minimized. Therefore, the comparative statics of such firms in "long-run" equilibrium can be analyzed through the behavioral postulate

$$\text{(1)} \qquad \text{minimize } AC = \sum_{i=1}^{n} p_i x_i / y$$

In other words, identical competitive firms in the long run behave "as if" they were average cost minimizers. Differentiation of AC with respect to x_i yields the first-order equation for long-run profit maximization:

$$\text{(2)} \qquad p_i - AC^* \cdot f_i^* = 0, \qquad i = 1, \ldots, n$$

where $f_i = \partial f / \partial x_i$ and the * denotes equilibrium (wealth maximizing) values. Differentiation of equation (2) again with respect to x_i yields the matrix of cross-partials of AC, (AC_{ij}). It is easily verified that

$$\text{(3)} \qquad AC_{ij} = -y f_{ij}, \quad i, j = 1, \ldots, n$$

A sufficient condition for AC to have a local minimum is that, together with equations (2), (AC_{ij}) be positive definite, which we assume to be the case. From equations (3) this implies that the matrix of cross-partials of the production function is negative definite, i.e., locally strictly concave. Hence the second-order conditions for this model are identical to those obtained in the standard

* University of Washington, Seattle.

[1] See Lowell Bassett and Thomas Borcherding (1970a, b, c), C. E. Ferguson and Thomas Saving, Paul Meyer (1967, 1968).

[2] Alternatively, the ensuing analysis can be interpreted as corresponding to the steady state of a firm which has positive adjustment costs.

[3] Equality of the min AC levels can always be achieved through the artifact of capitalizing the rents to factors specific to the firm into the firms' AC functions, assuming that markets exist for these specialized factors and that other firms value these factors equally with the present owners. These "rents" are then in fact opportunity costs to the firm. However, this device seems largely devoid of empirical significance in that this procedure is relevant only to the decision to remain in or leave the industry, but not to the level of production, since the marginal cost curves are unaffected by such capitalization of rents.

(output price constant) models (see for example, Samuelson (1947)).

The conditions (2) for average cost minimization (long-run profit maximization) say that the firm should employ factors until the wage equals the *average cost* times the marginal product. This is in conformance with the usual theory: In long-run equilibrium, output price equals the minimum average cost, and hence conditions (2) are just the same rules as the traditional setting of the value of marginal product of each factor equal to the respective wage.

The negative definiteness of (f_{ij}) guarantees that the Jacobian matrix associated with equations (2) is nonzero, and those equations can therefore be solved, in principle, for the long-run equilibrium factor demand curves

$$(4) \quad x_i^* = h_i(p_1, \ldots, p_n), \quad i = 1, \ldots, n$$

These demand curves are homogeneous of degree zero in the factor prices, since multiplication of those prices by any scalar leaves equations (2) (and the second-order conditions) unchanged.

We shall have occasion to discuss the "relative" factor demand relations

$$(5) \quad x_i^*/y^* = g_i(p_1, \ldots, p_n), \quad i=1, \ldots, n$$

These variables, the inverses of the average factor products, play a crucial role in the comparative statics of the firm in long-run equilibrium. The functions g_i are also homogeneous of degree zero, since if x_i^* is unchanged by scalar multiplication of the prices, y^* must also be unchanged.

The comparative statics of the firm in long-run equilibrium can be derived by the usual method of differentiating equations (2) with respect to the parameters involved, in this case, p_i. We shall not adopt this approach, however. Instead, a technique first suggested by Samuelson (1965) and developed more fully in my 1974 paper will be used. However, from general considerations concerning the form of equations (2), it is apparent that it *cannot* be proved that for the general case $\partial x_i^*/\partial p_i<0$, or that $\partial x_i^*/\partial p_j=\partial x_j^*/\partial p_i$, two results which are valid for the traditional case of fixed output price (see Samuelson (1947)). The appearance of each p_i in every equilibrium equation (2) precludes, a priori, the possibilities of deriving qualitative restrictions on the above partial derivatives from the maximization hypothesis alone. It will be shown, however, that qualitative restrictions are indeed available for the variables x_i^*/y^*, $i=1, \ldots, n$.

The indirect average cost function $AC^*(P)$ is derived by substituting the demand relations (4) into AC, i.e.,

$$\begin{aligned}(6) \quad AC^* &= \sum p_i x_i^*/y^* \\ &= \sum p_i h_i(P)/f(h_1(P), \ldots, h_n(P)) \\ &= \sum p_i g_i(P)\end{aligned}$$

where $P=(p_1, \ldots, p_n)$ is the vector of input prices. Consider now the "primal-dual" function $z(X, P) = AC^*(P) - AC(X, P)$, where $z(X, P)$ is considered to be a function of the $2n$ independent variables (X, P).[4] Since AC^* is the *minimum* value of average cost for any given price vector, and AC is any other average cost for a nonoptimal input mix, the difference, $z=AC^*-AC$, must always be less than or equal to zero, and must achieve a maximum value (of zero) with respect to X and P at $X=X^*$, i.e., when equations (2) or (4) hold. Therefore, the following first-order conditions for a maximum of $z(X, P)$ necessarily hold at $X=X^*$:

$$\begin{aligned}(7) \quad \partial z/\partial x_i &= -\partial AC/\partial x_i \\ &= (1/y)(p_i - (AC)f_i) = 0, \\ &\qquad i = 1, \ldots, n\end{aligned}$$

$$\begin{aligned}(8) \quad \partial z/\partial p_i &= \partial AC^*/\partial p_i - \partial AC/\partial p_i \\ &= \partial AC^*/\partial p_i - x_i/y = 0, \\ &\qquad i = 1, \ldots, n\end{aligned}$$

Equations (7) are merely equation (2) again. Equations (8) are a statement of the Viner-Wong envelope theorem (see Samuelson (1947)) for this model. For equilibrium values, the rate of change of the objective function with respect to a parameter, letting

[4] This is a departure from Samuelson's (1965) analysis, in which he treats first X, and then P alone, as independent variables.

all the primary variables adjust, is equal to the rate of change letting no variables adjust. In particular,

$$\partial AC^*/\partial p_i = \overset{*}{x_i}/\overset{*}{y} + \sum_{j=1}^{n} (\partial AC/\partial x_j)(\partial h_j/\partial p_i)$$
$$= \overset{*}{x_i}/\overset{*}{y}$$

since $\partial AC/\partial x_j = 0, j=1, \ldots, n$, by equation (2).[5] Hence equations (8) can be written

$$(9) \qquad \partial z/\partial p_i = \overset{*}{x_i}/\overset{*}{y} - x_i/y = 0, \qquad i = 1, \ldots, n$$

Since $z(X, P)$ has a maximum at $X = X^*$, the matrix of cross-partials of z must be negative semidefinite, and the principal minors of this matrix must be of sign $(-1)_k$, $k=1, \ldots, n$.[6] Among these principal minors are the principal minors of the submatrix of terms of the form[7]

$$\partial^2 z/\partial p_i \partial p_j = \partial(\overset{*}{x_i}/\overset{*}{y} - x_i/y)/\partial p_j$$
$$= \partial(\overset{*}{x_i}/\overset{*}{y})/\partial p_j$$

Since this is the crux of the theory to be presented, the result is stated as a formal theorem:

THEOREM 1: *The qualitative comparative statics of firms of the type discussed in the text above, i.e., those whose wealth-maximizing goal results in average cost minimization due, for example, to zero entry and adjustment cost conditions in the industry, can be summarized in the statement that the matrix of cross-partial derivatives of the type $\partial(x_i^*/y^*)/\partial p_j$, $i, j = 1, \ldots, n$ is symmetric and negative semidefinite.*

The symmetry of this matrix follows from $\partial(x_i^*/y^*)/\partial p_j = \partial^2 z/\partial p_i \partial p_j$ and the mathematical theorem that cross-partial derivatives are invariant with respect to the order of differentiation. This symmetry immediately yields the following reciprocity relation.[8] (To alleviate notational clutter, the asterisks will now be dropped. All variables are assumed to pertain to equilibrium values.)

COROLLARY 1: $\partial(x_i/y)/\partial p_j = \partial(x_j/y)/\partial p_i$, $i, j = 1, \ldots, n$

Whereas it is not in general true that $\partial x_i/\partial p_j = \partial x_j/\partial p_i$, the "relative" factor demands x_i/y do exhibit this curious reciprocity.[9]

COROLLARY 2: *The relative factor demand curves are downward sloping, i.e.,*

$$(10) \qquad \partial(x_i/y)/\partial p_i < 0, \qquad i = 1, \ldots, n$$

This is an immediate consequence of negative semidefiniteness: the diagonal elements of the matrix $\partial(x_i/y)/\partial p_j$ must all be negative.[10,11]

[5] Ferguson and Saving were apparently unaware that Samuelson's envelope theorem provided a one-line proof that $\partial AC^*/\partial p_i = x_i^*/y^*$. See their Note 1, Appendix, p. 780.

[6] See Samuelson (1947, Appendix A). More precisely, since only negative *semi*definiteness is implied by maximization, the principal minors have sign $(-1)^k$ or zero, $k=1, \ldots, n$. The determinant of the whole matrix $(\partial^2 z/\partial p_i \partial p_j)$ must be zero due to the invariance of the maximum position to a proportionate change in all prices.

[7] The "variables" x_i and hence y are independent variables in this partial differentiation; only the equilibrium values x_i^* and hence y^* are functions of the factor prices.

[8] For a more complete discussion of the derivation of reciprocity conditions see my 1972 paper.

[9] This reciprocity relation is derivable by conventional methods, of course: Letting D be the determinant of the matrix of cross-partial derivatives of AC with respect to x_i, with D_{ij} the (signed) cofactor of the element in the ith row, jth column,

$$\begin{aligned}\partial(x_i/y)/\partial p_j &= (1/y^2)\left[(y\partial x_i/\partial p_j) - x_i \sum_{k=1}^{n} f_k(\partial x_k/\partial p_j)\right]\\ &= (1/y^2 D)\left\{y\left[-D_{ji} + x_j/y \sum_{k=1}^{n} f_k D_{ki}\right]\right.\\ &\quad \left. - x_i\left[\sum_{k=1}^{n} f_k\left(-D_{jk} + (x_j/y)\sum_{h=1}^{n} f_h D_{hk}\right)\right]\right\}\\ &= (1/y^2 D)\left\{-yD_{ji} + x_j \sum_{k=1}^{n} f_k D_{ki}\right.\\ &\quad \left. + x_i \sum_{k=1}^{n} f_k D_{kj} - (x_i x_j/y) \sum_{h=1}^{n}\sum_{k=1}^{n} f_h D_{hk}\right\}\end{aligned}$$

an expression which is symmetric in i and j. The reason for deriving this result by the new methodology offered in the text is obvious.

[10] Strictly speaking, only a weak inequality is implied by maximization, i.e., the relative factor demands can-

An alternative interpretation of Corollary 2 is that with a rise (fall) in a factor price, the average product of that factor must eventually rise (fall) also. This is a refutable hypothesis for which data might actually exist: average quantities are relatively easy to come by. This proposition is not to be construed, however, as saying that the average product curves are upward sloping. Such curves are drawn holding the other factors fixed; the above interpretation of Corollary 2 concerns holding other factor *prices* constant and letting the other input levels adjust according to equations (4) or (5).

Using the quotient rule for differentiation, equation (10) yields $y\partial x_i/\partial p_i - x_i \partial y/\partial p_i < 0$. Hence, Corollary 2 expressed in elasticities is

COROLLARY 3:

$$(11)\quad (p_i/x_i)\partial x_i/\partial p_i < (p_i/y_i)\partial y_i/\partial p_i, \quad i = 1, \ldots, n$$

That is, the restriction on the elasticities of the factor demand curves implied by average cost minimization is that the elasticity of factor demand must be less than the output elasticity with respect to that same factor's price. Hence only in the case where the firm size (measured by output) remains unchanged or moves oppositely to a factor price change can one be sure that factor demand curves (*for the firm*) are downward sloping in the long run, i.e., after the firm has responded to output price changes.

From the alternation in sign of the principal minors of $\partial(x_i/y)/\partial p_j$, for the 2x2 case, one derives $[\partial(x_i/y)/\partial p_i][\partial(x_j/y)/\partial p_j] > [\partial(x_i/y/\partial p_j]^2$. It is difficult to provide an economic interpretation to these expressions; they are another example of how, in maximizing systems, "own effects" dominate "cross-effects," such as J. R. Hicks found in the theory of the consumer. These results and those of Hicks follow from the same mathematical theorem on negative semidefinite matrices.

COROLLARY 4: *If additional restrictions are placed on the acquisition of factor inputs (consistent with the original equilibrium), for example, point or outright rationing, the relative factor demands x_i/y, $i = 1, \ldots, n$ become more inelastic.*

This result follows from the general envelope properties developed by the author (1971). Also, the absolute cross elasticities of the relative factor demands can be expected to increase with direct rationing.

COROLLARY 5: *For the case of only two factors of production, $\partial(x_1/x_2)/\partial p_1 < 0$, $\partial(x_2/x_1)/\partial p_2 < 0$. For example if a firm employs "labor" and "capital" only, the capital-labor ratio must increase with a rise in wages.*

PROOF:

From the homogeneity of the relative factor demand curves (5), Euler's theorem yields $(\partial(x_2/y)/\partial p_1)p_1 + (\partial(x_2/y)/\partial p_2)p_2 \equiv 0$. The second term is necessarily negative; hence the first term is positive. Hence

$$\begin{aligned}\partial(x_1/x_2)/\partial p_1 &= \partial[(x_1/y)/(x_2/y)]/\partial p_1 \\ &= (y/x_2)^2[(x_2/y)\partial(x_1/y)/\partial p_1 \\ &\quad - (x_1/y)\partial(x_2/y)/\partial p_1] < 0\end{aligned}$$

In like manner, $\partial(x_2/x_1)/\partial p_2 < 0$.

We now analyze the response of the output price and output level of the firm to changes in factor prices.

THEOREM 2: *The output price elasticity with respect to a factor price, η_{pp_i} equals the share of total cost going to that factor, κ_i, where $p = \min AC$.*

PROOF:

Using equations (8), $\eta_{pp_i} = (p_i/p)(\partial p/\partial p_i) = p_i x_i/py = \kappa_i$.

not be upward sloping. We assume sufficient regularity of the production function so as to guarantee *negative* slopes.

[11] This result was derived algebraically by Bassett and Borcherding (1970b). A much simpler proof is: Let X^0, X^1 be the input vectors that minimize AC at factor price vectors P^0, P^1, respectively. Then (using an obvious scalar product notation) $P^0X^0/y^0 \leq P^0X^1/y^1$ and $P^1X^1/y^0 \leq P^1X^0/y^0$. Adding these two inequalities and rearranging terms yields $(P^0-P^1)(X^0/y^0 - X^1/y^1) \leq 0$. If only one factor price, say p_i changes, this expression becomes $\Delta p_i \cdot \Delta(x_i/y) \leq 0$. If the minimum point is assumed to be unique the weak inequality can be dropped to yield the stronger result of downward sloping relative factor demand curves.

THEOREM 3: $\partial y/\partial p_i=0$, $i=1, \ldots, n$ *if, and only if*, $\partial x_i/\partial p_j=\partial x_j/\partial p_i$.

PROOF:

Since $\partial(x_i/y)\partial p_j=\partial(x_j/y)/\partial p_i$ always, then using the quotient rule for differentiation, $y\partial x_i/\partial p_j-x_i\partial y/\partial p_j=y\partial x_j/\partial p_i-x_j\partial y/\partial p_i$. If $\partial y/\partial p_i=0$, $i=1, \ldots, n$, then $\partial x_i/\partial p_j=\partial x_j/\partial p_i$. Conversely, if $\partial x_i/\partial p_j=\partial x_j/\partial p_i$,

$$\begin{aligned}\partial y/\partial p_i &= \sum_{k=1}^{n} f_k\partial x_k/\partial p_i \\ &= 1/AC\sum_{k=1}^{n}(\partial x_k/\partial p_i)p_k \\ &= 1/AC\sum_{k=1}^{n}(\partial x_i/\partial p_k)p_k = 0\end{aligned}$$

by homogeneity of the demand functions (4). Hence the reciprocity relations which hold always in the short run (output price held constant), profit maximizing models are valid in the long run when, and only when, output is invariant to changes in factor prices.

THEOREM 4. *When a factor price increases, the resulting scale adjustment of the firm is to increase (decrease) output if the output elasticity is less than (greater than) unity; the output level remains unchanged if the output elasticity is unity. That is,* $\partial y/\partial p_i \gtrless 0$ *as* $\eta_{iy_0}\lesseqgtr 1$, *where* $\eta_{iy_0}=(y_0/x_i)(\partial x_i/\partial y_0)$.

Remark: Ferguson and Saving develop a similar theorem to this one, using the concept of an expenditure elasticity of factor i, or $\eta_{ic}=(c/x_i)\partial x_i/\partial c$; where $c=\sum_{i=1}^{n} p_ix_i$. Although the expressions developed by Ferguson and Saving in terms of the expenditure elasticities are ultimately valid if interpreted correctly, the method of derivation is certainly invalid and can lead to errors. The problem is that the variable c is treated as a parameter, as an independent variable in the demand functions for x_i, whereas in the model used by Ferguson and Saving (and here), c is clearly a dependent variable, being a function of x_i (and ultimately p_i, through the demand relations (4)). On a mathematical level, then, the derivative $\partial x_i/\partial c$ simply makes no sense. In terms of economics, there is confusion with regard to the problem of *ceteris paribus*. It is impossible to autonomously vary total expenditure without varying some factor price also. In order to rescue the concept of an expenditure elasticity, one must have in mind a model in which total expenditure is held constant in the factor demand curves, in a manner analogous to the utility maximization problem in consumer theory.

However, neither is "output" a parameter in the above model. The partial derivative $\partial x_i/\partial y_0$ has no meaning in the context of a minimization of AC scheme, since it is also impossible to vary output in this model unless some factor price changes. The output elasticities refer to the characteristics of the production function, as do the average and marginal cost functions (for a given factor market structure). These concepts are derived from the (total) cost minimization scheme, minimize $\sum p_ix_i$, subject to $f(x_1, \ldots, x_n)=y_0$, where y_0 is a parametrically fixed level of output. The resulting demand curves $x_i=\phi(p_1, \ldots, p_n, y_0)$ allow differentiation with respect to output, holding factor prices constant. The ensuing analysis incorporates these changes when necessary.

PROOF:

In Theorem 4, the minimum AC point occurs when $MC(P, y_0)=AC(P, y_0)$.

This equation can be used to define $y_0=y_0(p_1, \ldots, p_n)$. Hence differentiating implicitly with respect to p_i yields

$$\begin{aligned}(12)\quad \partial MC/\partial p_i &+ (\partial MC/\partial y_0)(\partial y_0/\partial p_i) \\ &= \partial AC/\partial p_i + (\partial AC/\partial y_0)(\partial y_0/\partial p_i)\end{aligned}$$

However, at the min AC point, $\partial AC/\partial y_0=0$. Hence[12]

$$\begin{aligned}(13)\quad \partial y_0/\partial p_i &= (1/(\partial MC/\partial y_0) \\ &\quad\cdot[\partial AC/\partial p_i-\partial MC/\partial p_i]\end{aligned}$$

However,[13] $\partial AC/\partial p_i=x_i/y$, $\partial MC/\partial p_i=\partial x_i/\partial y_0$. Factoring out x_i/y_0 yields

$$\begin{aligned}(14)\quad \partial y_0/\partial p_i &= (x_i/y_0)(1/(\partial MC\partial y_0)) \\ &\quad\cdot(1-\eta_{iy_0})\end{aligned}$$

[12] This provides an easy proof that $\partial y_0/\partial p_i \gtrless 0$ as $\partial AC/\partial p_i \gtrless \partial MC/\partial p_i$. Output increases or decreases as the AC curve shifts more or less than the MC curve.

[13] See Samuelson (1947, equation (46), p. 66).

from which Theorem 4 follows. (The slope of the marginal cost curve, $\partial MC/\partial y_0$ must be positive.) In terms of elasticities, equation (14) can be written

$$\eta_{y_0 p_i} = (\kappa_i/\epsilon_{MC,y_0})(1 - \eta_{iy_0}) \tag{15}$$

where $\eta_{y_0 p_i} = (p_i/y_0)\partial y_0/\partial p_i$, the output elasticity with respect to a factor price change, κ_i = share of total cost spent on factor i, and ϵ_{MC,y_0} = elasticity of marginal cost curve.

THEOREM 5: *If the production function is homothetic,* $\partial y_0/\partial p_i = 0$, $i = 1, \ldots, n$. *Conversely, if* $\partial y_0/\partial p_i = 0$, $i = 1, \ldots, n$, *then the function is locally homothetic.*[14]

PROOF:

Along any isoquant, $\sum_{i=1}^{n} f_i(\partial x_i/\partial y_0) = 1$. At the min AC point, $f_i = p_i/AC$, hence in terms of elasticities, at min AC,

$$\sum_{i=1}^{n} \kappa_i \eta_{iy_0} = 1 \tag{16}$$

If $f(x_1, \ldots, x_n)$ is homothetic, then all the output elasticities must be equal, since the expansion paths are straight lines through the origin. Since the shares must sum to one, by equation (16), the output elasticities are all unity and hence $\partial y_0/\partial p_i = 0$, $i = 1, \ldots, n$ by Theorem 4. Conversely, if $\partial y_0/\partial p_i = 0$, $i = 1, \ldots, n$, then the output elasticities are all unity. Since they are all equal, the expansion path at that point must lie along a line emanating from the origin; hence the production function is locally homothetic.

COROLLARY: *If the production function is locally homothetic, then the ordinary factor demand curves are downward sloping.*

PROOF:

For homothetic production functions, $\partial y_0/\partial p_i = 0$, $1, \ldots, n$, hence by equation (11), $\partial x_i/\partial p_i < 0$.

Theorems 3, 4, and 5 say that the following four statements are equivalent for firms at their long-run min AC levels:

(i) $\partial y/\partial p_i = 0 \qquad i = 1, \ldots, n$

(ii) $\partial x_i/\partial p_j = \partial x_j/\partial p_i, \qquad i, j = 1, \ldots, n$

(iii) $\eta_{iy_0} = 1 \qquad i = 1, \ldots, n$

(iv) The production function is locally homothetic.

Consider a particular factor price vector P and suppose that min AC occurs at some output level y_0. This min AC level is a feature of how "compressed" the isoquants are for equal output changes. Under homotheticity, the isoquants are all radial blow-ups of each other; hence it is impossible for the isoquants to be compressed differently along different rays from the origin. Geometrically, then, the result that the min AC output level is invariant to factor price changes for homothetic production functions is clear: for such production functions there is a min AC *isoquant* based solely on the technological characteristics of the production function, i.e., the relative "compressions" of the isoquants.

II. Nonidentical Firms and Industry Responses

The above analysis pertains only to the actions of individual firms, all of which have identical minimum AC levels. Important questions remain as to the response of input and output variables on the industry level when the parameters facing all firms change. We would like to know, for example, whether it is possible to derive downward sloping industry factor demands from the type of firm behavior analyzed above. One result is immediate. Since by Theorem 2 every firm's average cost must increase when a factor price increases, the industry's output price must increase resulting in a fall in industry output. This fall in industry output does not, however, translate into a fall in each firm's output. With nonidentical firms, some firms may increase in size in accordance with Theorem 4, if that factor is output-inelastic, and vice versa for output-elastic factors. These output effects will determine whether an individual firm will hire more or less of the factor whose price increased.[15]

[14] See also Shephard.

[15] Paradoxically, an increase in factor price can stimulate entry into the industry, if the factor is sufficiently output-elastic to outweigh the negative elasticity of industry demand.

The question remains as to the compatibility of the "as if" assumption of minimization of average cost and the assumption of nonidentical firms. Is it possible for firms with different production functions to have identical min AC levels? Equivalently, is it possible for the ratios x_i/y to vary among such firms in long-run equilibrium? The answer must in general be in the negative. For suppose the ratios x_i/y were different for different firms. Then if p_i changed, the AC curves for each firm would shift by different amounts, destroying the equality of min AC levels. Hence in order for all firms to have the same min AC level *for all sets of factor prices*, the production functions of the firms must all be identical except for scale; i.e., the firms are virtually indistinguishable. Moreover, which firms enter or leave such industries is a matter not capable of being analyzed by this or similar models; however, that question is probably unimportant for our present purposes.

When the production functions for each firm are different, the industry supply curve will be upward sloping. In that case, only the marginal firm (that firm in the industry with the highest min AC) will operate at the point of min AC, even in the long run. Factors specific to the intramarginal firms will earn rents. If some factor price p_i increases, then the industry output price will increase at the rate x_i/y given by the marginal firm. Each firm's AC curve, however, will shift by different amounts. Firms whose shares of revenue going to that factor are greater than κ_i for the marginal firm will experience a decline in rents, with the opposite holding for firms whose shares to that factor are less than that share in the marginal firm. There is no way to know a priori whether firms with the largest increases in AC will leave the industry first because these firms might also be the ones earning the largest rents. Because of this, it is not possible to identify how the total use of the factors in question by the industry will respond to a change in that factor's price. If a factor price p_i increases, the firms that are heavy users of x_i may expand relative to the other firms in the industry producing an upward sloping industry factor demand curve.[16] Whether or not this situation is empirically relevant, it appears that downward sloping factor demands for an industry composed of nonidentical firms will have to remain an asserted rather than a derived result. For the case of identical firms, of course, the industry's factor demands must be downward sloping as has been shown, for example, by R. G. D. Allen, Bassett and Borcherding (1970b). The result follows by noting that the common ratio x_i/y declines for each firm, and by simple addition (since the ratios are equal) since total industry output must decline, total factor utilization must decline even more.

III. Concluding Remarks

The purpose of this paper was to derive meaningful theorems about firms in their long-run competitive equilibrium. The model used in this paper allows this to be done in a relatively simple manner for industries composed of identical firms, without resort to complicated manipulations of bordered Hessian determinants as were used by Ferguson and Saving, as well as other writers. Through this greater simplicity, the fundamental nature of the relative factor demands x_i/y (the inverses of the average products) was revealed. For these variables, the usual downward sloping factor demand curves and reciprocity relations associated with production theory are valid. Only in the case of homothetic production functions are these relations also valid for the ordinary factor demand curves.

[16] The intramarginal firms will expand or contract according to the difference in the shifts of the MC curve of that firm and the AC curve of the marginal firm, the latter of which dictates the industry output price.

REFERENCES

R. G. D. Allen, *Mathematical Analysis for Economists*, New York 1964.

L. R. Bassett and T. E. Borcherding, (1970a) "The Firm, the Industry, and the Long-Run Demand for Factors of Production," *Canadian J. Econ.*, Feb. 1970, *3*, 140–44.

——— and ———, (1970b) "Industry Factor Demand," *Western Econ. J.*, Sept. 1970, *8*, 259–61.

——— and ———, (1970c) "The Relationship Between Firm Size and Factor Price," *Quart. J. Econ.*, Aug. 1970, *84*, 518–22.

D. V. T. Bear, "Inferior Inputs and the Theory of the Firm," *J. Polit. Econ.*, June 1965, *73*, 287–89.

C. E. Ferguson and T. R. Saving, "Long-Run Scale Adjustments of a Perfectly Competitive Firm and Industry," *Amer. Econ. Rev.*, Dec. 1969, *59*, 774–83.

J. R. Hicks, *Value and Capital*, Oxford 1946.

P. A. Meyer, "A Paradox on Profits and Factor Prices," *Amer. Econ. Rev.*, June 1967, *57*, 535–40.

———, "A Paradox on Profits and Factor Prices: Reply," *Amer. Econ. Rev.*, Sept. 1968, *58*, 923–30.

P. A. Samuelson, *Foundations of Economic Analysis*, Cambridge, Mass. 1947.

———, "Using Full Duality to Show that Simultaneously Additive Direct and Indirect Utilities Implies Unitary Price Elasticity of Demand," *Econometrica*, Oct. 1965, *33*, 781–96.

R. W. Shephard, *The Theory of Cost and Production Functions*, Princeton 1970.

E. Silberberg, "The Le Chatelier Principle as a Corollary to a Generalized Envelope Theorem," *J. Econ. Theory*, June 1971, *3*, 146–55.

———, "Reciprocity and Duality," *Western Econ. J.*, Mar. 1972, *10*, 95–100.

———, "A Revision of Comparative Statics Methodology in Economics," *J. Econ. Theory*, Feb. 1974, *7*, 159–72.

Relative Efficiency in Wheat Production in the Indian Punjab

By SURJIT S. SIDHU*

In recent contributions to this *Review*, Lawrence Lau and Pan Yotopoulos (L-Y) applied the profit function concept to the analysis of relative efficiency of Indian agriculture. They developed an operational model to measure and compare *economic efficiency* and its components of *technical efficiency* and *price* (or allocative) *efficiency* for groups of firms. By comparing the actual profit functions of small and large farms at given output and input prices and fixed quantities of land and capital, they found that smaller farms had higher profits (total revenue minus the total cost of the variable factors of production, in this case labor) than larger farms within the range of output studied and hence were economically more efficient. Further, they were able to show that the relative economic superiority of small farms was due to their technical efficiency since both types of farms were price efficient. Their results also indicate constant returns to scale in Indian agriculture. Both these findings have far-reaching implications for the optimal allocative structure of Indian agriculture.

In this paper, the L-Y model is confronted with new and recent data for wheat farms[1] in the Indian Punjab. My results run counter to their findings in that I do not find any differences in the economic efficiency (or its components of technical efficiency and price efficiency) of small and large wheat farms. Using their model, I also compare the economic performance of old Indian wheat varieties with Mexican varieties, and tractor-operated with non-tractor-operated wheat farms. The last mentioned two comparisons have considerable relevance in the context of the "green revolution" and the absorption of a rapidly growing labor force in India and other *LDC*s. Section I of the present paper establishes a link between my estimation procedure and the L-Y model and briefly describes the data and the variables. Section II provides empirical estimates, derives the implications of these results, and compares them with those of L-Y. Section III summarizes my conclusions.

I. Estimating Procedure and the Data

In order to study relative economic efficiency and its components of technical efficiency and price efficiency, the profit function formulation used by Lau and Yotopoulos seems to be an ideal tool. Since this paper is an extension of their work, it is unnecessary to reproduce the model and the related derivations. I provide a brief description of the data and the variables as they relate to the estimating equations.

To start with, let the wheat production function be written as:

$$(1) \qquad Y = F(N; L, K)$$

where Y is output, N is the variable input labor, and L and K are the fixed inputs of land and capital, respectively. Following the L-Y papers, the estimating equations for the Cobb-Douglas case of this production function are presented in Tables 2, 3, and 4.

For statistical specification I assume additive errors with zero expectation and finite variance for each of the two equations. The covariance of the errors of the two equations for the same farm may not be zero but the covariances of the errors of either equation

* Assistant professor, department of agricultural and applied economics, University of Minnesota. This paper is based upon my Ph.D. thesis at the University of Minnesota. Acknowledgement is made for financial support from the Rockefeller Foundation and Economic Development Center, University of Minnesota. I would like to thank Lee R. Martin, Vernon R. Ruttan, Willis L. Peterson, Martin E. Abel, and Hans P. Binswanger for many helpful suggestions and discussions.

[1] The Punjab farms are multi-enterprise farms. This investigation deals only with wheat, not all farm enterprises.

TABLE 1—BRIEF SUMMARY OF THE SAMPLES AND DATA

Sample	Geographic Coverage	Villages Included	Farms	Crop Year	Wheat Type	Observations Available
1) Ferozepur	District-Ferozepur	15	150	1967–68	New	105
				1967–68	Old	132
				1968–69	New	144
2) Tractor Cultivation	Punjab	19	304	1969–70	New	287
3) Regionally Stratified	Punjab	7	128	1970–71	New	128

Sources: 1) Directorate of Economics and Statistics, Ministry of Food and Agriculture, Government of India; 2) The Economic Adviser, Government of Punjab; 3) I was responsible for the design and supervision of data collection work for this sample.

corresponding to different farms are assumed to be zero. With these assumptions an asymptotically efficient method of estimation as proposed by Arnold Zellner is used[2] to estimate jointly the parameters of the two equations. A brief description of the variables and the notation used is as follows:

Y = physical output of wheat measured in quintals per farm including by-products.[3]

N = the labor input per farm used for wheat production measured in hours. It includes both family and hired labor.[4]

L = the land input measured as acres of wheat grown per farm.

K = a measure of flow of capital services going into wheat production per farm.[5]

p = the price of wheat per quintal as reported for each farm.

wN = the total wage bill in rupees for wheat production per farm.[6]

w = the hourly wage rate of labor. It is obtained simply by dividing the total wage bill wN by total labor input N.

P = the profit from wheat production: total revenue less total variable labor costs.

π = the profit function.

A brief summary of the samples and data used in this study is provided in Table 1. As compared to the group average data used by L-Y and most earlier Indian studies, I have been fortunate to have access to micro level primary data.

II. Empirical Results

A. *Old versus New Varieties of Wheat*

The first test for relative economic efficiency in wheat production in Punjab compares the economic efficiency of new varieties of wheat with the old varieties of wheat. For this purpose 1967–68 data are used from the Ferozepur Sample (see Table 1). The two equations (the profit function and the labor demand function) are estimated jointly using Zellner's method of estimation by imposing the restrictions that $\beta_1=\beta_1$ in the two equations and requiring that $\beta_2+\beta_3=1$, that is, assuming constant returns to scale.[7] These results are presented in Table 2. They indicate that the new wheats are economically

[2] This will also make our results comparable to those of L-Y.

[3] The by-products are converted into quintals of wheat by dividing the total value of by-products by wheat price.

[4] Child and female labor is converted into man equivalents by treating two children (or women) equal to one man.

[5] An hourly flow of services is derived for each durable input including capital in the form of livestock that the farm uses in wheat production. It includes depreciation charges, interest charges, and operating expenses. Depreciation schedules are based on the specific life of each input, but interest costs are estimated at a uniform interest rate of 10 percent per annum. The actual number of hours of use times the hourly flow of services of each durable input gives its total service flow. Aggregation of these asset-specific service flows plus the seed and fertilizer costs yields a measure of the capital services.

[6] Family labor services are valued as equivalent to those of the annual contract labor for each farm.

[7] This assumption is based on the earlier tests carried out in my dissertation and tests carried out in subsequent sections.

TABLE 2—RESULTS OF JOINT ESTIMATION OF COBB-DOUGLAS PROFIT FUNCTION AND LABOR DEMAND FUNCTION FOR WHEAT, 1967–68, PUNJAB, INDIA

Parameter	Estimated Coefficient	Standard Error
$\lambda=$	4.872	0.965
$\delta^N=$	0.485	0.129
$\beta_1=$	0.254	0.013[a]
$\beta_2=$	0.670	0.155
$\beta_3=$	0.330	0.155

Note: The estimating equations are:

$$\ln \pi = \lambda + \delta^N D^N + \beta_1 \ln w + \beta_2 \ln L + \beta_3 \ln K$$

$$-\frac{wN}{\pi} = \beta_1, \quad \text{where} \quad \lambda = \ln \overset{0}{A_*} + (1-\beta_1)\ln p$$

[a] In both equations.

more efficient compared to the old wheats by 48.50 percent. From $\lambda = \ln A_*^0 + (1-\beta_1) \ln p$ we evaluate A_*^0 by substituting the sample mean value of $\ln p$ for old wheat. Then we get A^0 the efficiency parameter in the Cobb-Douglas production function for old wheat, the computed value of which is 5.641. In the same way, from $\lambda = \ln A_*^0 + \delta^N + (1-\beta_1) \ln p$, we get A^N the efficiency parameter for new wheat = 8.166. Thus maintaining the *hypothesis of neutral technical shift*,[8] we find that the efficiency parameter for the new wheat production function is larger by 44.70 percent.

B. *Relative Efficiency*

There are different policy implications associated with each component of differences (technical efficiency or price efficiency) in economic efficiency of small and large farms. For example, the finding that small farms are more technical efficient and that both small and large farms are absolute price efficient could lead to the conclusion that small farms serve the national interest better (leaving aside the equity considerations). If we find that smaller farms are less price efficient, policies which improve market information for them may improve their allocative efficiency. Similar implications would follow if tractor-operated farms were more price efficient than non-tractor-operated farms. It is thus important to obtain knowledge of the source of differences (technical or price) in economic efficiency.

For the new varieties of wheat the estimation results using Zellner's method for each of the four years 1967–68 to 1970–71 and for the four-year combined data comparing small and large[9] wheat farms are presented in Table 3. For comparing tractor-operated and non-tractor-operated farms, the results employing data for Tractor Cultivation Sample 1969–70 (as described in Table 1), are presented in Table 4. In order to provide answers to the questions of relative efficiency posed above we carry out the following statistical tests:

1) The hypothesis of equal relative economic efficiency of small and large and tractor and nontractor wheat farms:

(i) H_o: $\delta^L = 0$

(ii) H_o: $\delta^T = 0$

2) The hypothesis of equal relative price efficiency with respect to labor demand of small and large and tractor and nontractor wheat farms:

(i) H_o: $\beta_1^L = \beta_1^S$

(ii) H_o: $\beta_1^T = \beta_1^{NT}$

3) The joint hypotheses of equal relative technical and price efficiency of small and large and tractor and nontractor wheat farms:

(i) H_o: $\delta^L = 0$ and $\beta_1^L = \beta_1^S$

(ii) H_o: $\delta^T = 0$ and $\beta_1^T = \beta_1^{NT}$

[8] In my dissertation I compared production functions for old and new varieties of wheat, small and large wheat farms, and tractor-operated and non-tractor-operated wheat farms and found that the differences in these production functions are only of the neutral type.

[9] In this study farms with more than 10 acres of wheat are defined as large farms and farms with 10 acres or less as small farms. This seems to be a realistic dividing line between large and small wheat farms for Punjab where the average farm size is 12.5 acres. (See Martin Billings and Arjan Singh, 1971.) Also it facilitates comparisons of our results with those of L-Y (1971, 1973) who also used this criterion for small and large farms.

TABLE 3—RESULTS OF JOINT ESTIMATION OF COBB-DOUGLAS PROFIT FUNCTION AND LABOR DEMAND FUNCTION FOR NEW WHEAT, PUNJAB, INDIA

	UOP Profit Function					Labor Demand Function		
	λ	δ^L	β_1	β_2	β_3	β_1^L	β_1^S	R^2
1967–68								
Single Equation	3.799	−0.141	0.107	0.614	0.487	−0.221	−0.289	0.923
Ordinary Least	*0.748*	*0.144*	*0.159*	*0.115*	*0.125*	*0.075*	*0.040*	
Squares (*OLS*)								
Unrestricted	3.433	−0.064	0.263	0.506	0.564	−0.221	−0.289	
	0.641	*0.137*	*0.136*	*0.098*	*0.107*	*0.075*	*0.040*	
1 Restriction[a]	3.446	−0.112	0.262	0.506	0.563	−0.274	−0.274	
	0.641	*0.123*	*0.136*	*0.098*	*0.107*	*0.035*	*0.035*	
2 Restrictions[b]	3.019	−0.138	−0.244	0.520	0.599	−0.244	−0.244	
	0.667	*0.131*	*0.034*	*0.104*	*0.113*	*0.034*	*0.034*	
3 Restrictions[c]	3.885	0.093	−0.236	0.537	0.462	−0.236	−0.236	
	0.636	*0.115*	*0.034*	*0.109*	*0.109*	*0.034*	*0.034*	
1968–69								
Single Equation	4.115	−0.041	−0.507	0.713	0.334	−0.406	−0.433	0.771
OLS	*0.994*	*0.160*	*0.207*	*0.179*	*0.170*	*0.065*	*0.059*	
Unrestricted	3.714	0.049	0.024	0.514	0.454	−0.406	−0.433	
	0.692	*0.133*	*0.144*	*0.124*	*0.118*	*0.065*	*0.059*	
1 Restrictions[a]	3.725	0.026	0.024	0.514	0.454	−0.421	−0.421	
	0.691	*0.111*	*0.144*	*0.124*	*0.118*	*0.043*	*0.043*	
2 Restrictions[b]	3.391	0.061	−0.381	0.477	0.495	−0.381	−0.381	
	0.673	*0.109*	*0.041*	*0.122*	*0.116*	*0.041*	*0.041*	
3 Restrictions[c]	3.309	0.015	−0.381	0.498	0.503	−0.381	−0.381	
	0.655	*0.070*	*0.041*	*0.114*	*0.114*	*0.041*	*0.041*	

Note: Zellner's method of estimation was used to obtain joint estimates of the *UOP* Profit Function and the Labor Demand Function.

[a] 1 Restriction: $\beta_1^L=\beta_1^S$.

[b] 2 Restrictions: $\beta_1^L=\beta_1$; $\beta_1^S=\beta_1$.

[c] 3 Restrictions: $\beta_1^L=\beta_1$; $\beta_2+\beta_3=1$; $\beta_1^S=\beta_1$.

The estimating equations for the four individual years are:

$$ln\ \pi = \lambda + \delta^L D^L + \beta_1 ln\ w + \beta_2 ln\ L + \beta_3 ln\ K$$

$$-\frac{wN}{\pi} = \beta_1^L D^L + \beta_1^S D^S$$

The estimating equations for the four-years' pooled data are:

$$ln\ \pi = ln\ A_*^S + \delta^L D^L + \sum_{i=1}^{3} \delta_i D_i + \beta_1\ ln\ w + \beta_2\ ln\ L + \beta_3\ ln\ K$$

$$-\frac{wN}{\pi} = \beta_1^L D^L + \beta_1^S D^S$$

where

π = profit (total receipts less wage bill)
w = money wage rate
D^L = dummy variable taking the value of one if wheat area is greater than ten acres and zero otherwise
D^S = dummy variable taking the value of one if wheat area is less than ten acres and zero otherwise
D_i = three year dummy variables taking the value of one for 1968–69, 1969–70, and 1970–71, respectively, and zero otherwise
N = labor in hours per farm used in wheat production
L = land in acres used for producing wheat
K = total costs of capital services for wheat per farm

Asymptotic standard errors are in italics.

TABLE 3 (*Continued*)

	UOP Profit Function					Labor Demand Function		
	λ	δ^L	β_1	β_2	β_3	β_1^L	β_1^S	R^2
1969–70								
Single Equation	4.651	0.093	−0.278	0.740	0.259	−0.501	−0.449	0.776
OLS	*0.477*	*0.108*	*0.124*	*0.098*	*0.082*	*0.153*	*0.204*	
Unrestricted	4.748	0.136	−0.058	0.714	0.260	−0.501	−0.440	
	0.411	*0.098*	*0.106*	*0.085*	*0.070*	*0.153*	*0.204*	
1 Restriction[a]	4.744	0.142	−0.058	0.714	0.260	−0.482	−0.482	
	0.410	*0.093*	*0.016*	*0.085*	*0.070*	*0.122*	*0.122*	
2 Restrictions[b]	4.744	0.142	−0.248	0.716	0.256	−0.248	−0.248	
	0.418	*0.094*	*0.081*	*0.086*	*0.072*	*0.081*	*0.081*	
3 Restrictions[c]	4.694	0.099	−0.247	0.742	0.257	−0.247	−0.247	
	0.408	*0.055*	*0.081*	*0.072*	*0.072*	*0.081*	*0.081*	
1970–71								
Single Equation	2.859	0.056	−0.481	0.477	0.581	−0.234	−0.304	
OLS	*0.641*	*0.110*	*0.189*	*0.131*	*0.112*	*0.051*	*0.048*	
Unrestricted	3.287	−0.048	−0.184	0.496	0.539	−0.254	−0.265	
	0.595	*0.104*	*0.176*	*0.121*	*0.103*	*0.051*	*0.046*	
1 Restriction[a]	3.291	−0.051	−0.184	0.496	0.539	−0.259	−0.259	
	0.594	*0.102*	*0.175*	*0.121*	*0.103*	*0.025*	*0.025*	
2 Restrictions[b]	3.386	−0.057	−0.255	0.512	0.523	−0.255	−0.255	
	0.581	*0.101*	*0.025*	*0.117*	*0.100*	*0.025*	*0.025*	
3 Restrictions[c]	3.438	−0.010	−0.254	0.477	0.523	−0.254	−0.254	
	0.576	*0.059*	*0.025*	*0.110*	*0.100*	*0.025*	*0.025*	

	UOP Profit Function								Labor Demand Function	
	$\ln A_*^S$	δ^L	δ_1	δ_2	δ_3	β_1	β_2	β_3	β_1^L	β_1^S
1967–68 to 1970–71										
Single Equation	4.405	−0.025	−0.411	−0.393	−0.242	−0.243	0.709	0.359	−0.411	−0.351
OLS	*0.334*	*0.059*	*0.068*	*0.063*	*0.071*	*0.079*	0.058	0.056	0.078	0.078
Unrestricted	4.479	−0.021	−0.384	−0.353	−0.241	−0.085	0.690	0.358	−0.412	−0.346
	0.301	*0.056*	*0.061*	*0.057*	*0.064*	*0.072*	*0.053*	*0.051*	*0.078*	*0.077*
1 Restriction[a]	4.475	−0.012	−0.384	−0.353	−0.240	−0.085	0.690	0.358	−0.379	−0.379
	0.301	*0.053*	*0.061*	*0.057*	*0.064*	*0.072*	*0.052*	*0.051*	*0.052*	*0.052*
2 Restrictions[b]	4.410	−0.015	−0.377	−0.347	−0.200	−0.279	0.700	0.358	−0.279	−0.279
	0.303	*0.054*	*0.062*	*0.057*	*0.063*	*0.042*	*0.053*	*0.051*	*0.042*	*0.042*
3 Restrictions[c]	4.568	0.075	−0.336	−0.305	−0.163	−0.271	0.663	0.337	−0.271	−0.271
	0.297	*0.038*	*0.060*	*0.054*	*0.061*	*0.042*	*0.050*	*0.050*	*0.042*	*0.042*

From Table 5 we see that none of these three hypotheses can be rejected at the 90 percent level of significance. Thus the hypotheses 1) that small and large wheat producing farms have equal relative *economic efficiency* and equal relative *price efficiency* and 2) that tractor-operated and non-tractor-operated wheat producing farms have equal relative *economic efficiency* and equal relative *price efficiency* are supported by these results. This implies that these farms also have equal *technical efficiency*. The view that small and large farmers have the same degree of economic motivation seems to hold. Because wheat is a dominant enterprise on these farms, one can argue that these conclusions would perhaps be equally applicable to all enterprises on these farms.

4a) Next maintaining the hypothesis of equal price efficiency in 2), we turn to the hypotheses of:

(i) Absolute price efficiency of large farms,

$$H_o:\ \beta_1^L = \beta_1$$

TABLE 4—RESULTS OF JOINT ESTIMATION OF COBB-DOUGLAS PROFIT FUNCTION AND LABOR DEMAND FUNCTION FOR NEW WHEAT, 1969–70, PUNJAB, INDIA

		Estimated Coefficients				
		Single Equation	Zellner's Method with Restrictions			
Function	Parameter	*OLS*	Un-restricted	1 Restriction[a]	2 Restrictions[b]	3 Restrictions[c]
UOP Profit Function	λ	4.830	4.778	4.794	4.811	4.934
		(0.501)	(0.433)	(0.433)	(0.441)	(0.398)
	δ^T	0.089	0.075	0.032	0.041	0.062
		(0.073)	(0.070)	(0.063)	(0.064)	(0.054)
	β_1	−0.286	−0.062	−0.064	−0.253	−0.256
		(0.124)	(0.107)	(0.107)	(0.081)	(0.081)
	β_2	0.790	0.785	0.785	0.788	0.779
		(0.083)	(0.071)	(0.071)	(0.073)	(0.072)
	β_3	0.224	0.241	0.241	0.235	0.221
		(0.085)	(0.073)	(0.073)	(0.075)	(0.072)
Labor Demand Function	β_1^T	−0.259	−0.259	−0.481	−0.252	−0.256
		(0.203)	(0.202)	(0.122)	(0.081)	(0.081)
	β_1^{NT}	−0.610	−0.610	−0.481	−0.252	−0.256
		(0.153)	(0.153)	(0.122)	(0.081)	(0.081)
	R^2	0.777				

Note: The estimating equations are:

$$\ln \pi = \lambda + \delta^T D^T + \beta_1 \ln w + \beta_2 \ln L + \beta_3 \ln K$$

$$-\frac{wN}{\pi} = \beta_1^T D^T + \beta_1^{NT} D^{NT}$$

where D^T is a dummy variable taking the value of one for farms owning a tractor and zero otherwise; and D^{NT} is a dummy variable taking the value of one for farms not owning a tractor (animal operated) and zero otherwise. Asymptotic standard errors are in parentheses.

[a] See fn. a, Table 3.
[b] See fn. b, Table 3.
[c] See fn. c, Table 3.

TABLE 5—TESTING OF STATISTICAL HYPOTHESES

Hypotheses		Computed F-Ratio and Degrees of Freedom				
Maintained	Tested	1967–68	1968–69	1969–70	1970–71	1967–68 to 1970–71
	1) $\delta^L=0$	0.22; (1,203)	0.14; (1,265)	1.93; (1,567)	0.23; (1,249)	0.15; (1,1302)
	2) $\beta^L=\beta^S$	0.64; (1,203)	0.10; (1,265)	0.04; (1,567)	0.02; (1,249)	0.34; (1,1302)
	3) $\delta^L=0$, $\beta^L=\beta^S$	0.73; (2,203)	0.08; (2,265)	1.20; (2,567)	0.14; (2,249)	0.20; (2,1302)
$\beta_1^L=\beta_1^S$	4) $\beta_1^L=\beta_1$	7.72; (2,203)	4.71; (2,265)	3.44; (2,567)	0.10; (2,249)	5.58; (2,1302)
$\beta_1^L=\beta_1^S$	5) $\beta_1^S=\beta_1$	7.72; (2,203)	4.71; (2,265)	3.44; (2,567)	0.10; (2,249)	5.58; (2,1302)
	6) $\beta_2+\beta_3=1$	839.81; (1,203)	373.61; (1,265)	384.94; (1,567)	306.41; (1,249)	1812.13; (1,1302)
	1) $\delta^T=0$			1.13; (1,567)		
	2) $\beta_1^T=\beta_1^{NT}$			1.92; (1,567)		
	3) $\delta^T=0$, $\beta_1^T=\beta_1^{NT}$			1.09; (2,567)		
$\beta_1^T=\beta_1^{NT}$	4) $\beta_1^T=\beta_1$			4.31; (2,567)		
$\beta_1^T=\beta_1^{NT}$	5) $\beta_1^{NT}=\beta_1$			4.31; (2,567)		
	6) $\beta_2+\beta_3=1$			914.14; (1,567)		

Note: Critical F-ratios are: $F_{0.10}$ $(1, \propto)=2.70$; $F_{0.05}$ $(1, \propto)=3.84$; $F_{0.10}$ $(1, \propto)=6.63$
$F_{0.10}$ $(2, \propto)=2.30$; $F_{0.05}$ $(2, \propto)=3.00$; $F_{0.01}$ $(2, \propto)=4.61$

(ii) Absolute price efficiency of small farms,

$$H_o: \beta_1^S = \beta_1$$

For the first two years 1967–68 and 1968–69 and for the four years pooled data, we reject these hypotheses at 99 percent level of significance and for the year 1969–70 at 95 percent level of significance. But, for the latest year 1970–71, we cannot reject these hypotheses at 90 percent level of significance. This means that during the years 1967–68, 1968–69, and 1969–70, both small and large farms were not in a state of equilibrium in the sense of equating the value of marginal product of labor to its wage rate. For the year 1970–71, however, we find that both small and large farms were in equilibrium, i.e., maximizing profits.

4b) Maintaining the hypothesis of equal price efficiency in 2), we also test the hypotheses of:

(i) Absolute price efficiency of tractor farms, $Ho: \beta_1^T = \beta_1$

(ii) Absolute price efficiency of nontractor farms, $Ho: \beta_1^{NT} = \beta_1$

The meaning of these tests is whether tractor and nontractor farms maximize profits by equating the value of marginal product of labor to its opportunity price. The null hypothesis is rejected in both cases. The conclusion is that both tractor and nontractor farms were not able to maximize profits during the year 1969–70. In light of the results for the hypothesis of equal relative price efficiency in statistical test 2), we conclude that, with respect to labor, tractor and nontractor farms were equally unsuccessful in their efforts to maximize profits by using the optimum amount of labor.

5) Lastly, we test the hypothesis of constant returns to scale in all factors of production:

$$H_o: \beta_2 + \beta_3 = 1$$

This hypothesis is rejected at the 99 percent level of significance in all cases. The sum $(\beta_2+\beta_3)>1$ for the years 1967–68, 1970–71, and for the four-year pooled data. But $(\beta_2+\beta_3)<1$ for the years 1968–69 and 1969–70. These differences from unity are quite small in either case. Also, perhaps slightly increasing returns for the years 1967–68 and 1970–71 resulted because a larger number of observations for these years were below the respective sample averages.

Important policy implications follow from these findings. The most substantive one is that policies with respect to land redistribution and ceilings on ownership of land can be based primarily on social and political considerations. Secondly, governmental policies with respect to pricing, supply of agricultural inputs, marketing facilities, provision of credit and extension services, etc. need not favor either large or small farms (or farms having tractors or without tractors) on the basis of their economic efficiency or its components of technical efficiency or price efficiency.

The results of statistical test 4) have interesting implications with respect to the profit-maximizing behavior (or rationality) of the wheat producers. They have a bearing on earlier price and allocative studies.[10] The results appear to indicate the existence of a disequilibrium between the profit-maximizing attempts and the actual results achieved by wheat producers. It is most likely created by a shift to the right in the labor demand function, resulting from the introduction of high-yielding wheats.[11] During the first three years, 1967–68, 1968–69, and 1969–70, producers were not in equilibrium in the sense of equating the marginal value product of labor to its opportunity cost. And during the last year 1970–71, we cannot reject the hypotheses of absolute price efficiency. That is, we find that producers on the average (both small and large) were able to equate the marginal value product of labor to its going opportunity cost. These results seem to suggest that in a changing agriculture one should

[10] See for example David Hopper, A. M. Khusro, Theodore Schultz, G. S. Sahota and L-Y (1971, 1973).

[11] Results reported in my dissertation indicate that the per acre factor demand functions shifted to the right by 25 percent as a result of the introduction of Mexican wheat varieties in Punjab.

expect the existence of a fair amount of inefficiency in the labor market but that producers do seem to react to overcome the existence of a disequilibrium.

C. *Comparison with Findings by Lau and Yotopoulos*

We provide two brief comparisons of our results with the research of L-Y (1971, 1973) regarding relative efficiency in Indian agriculture.

Estimates for the Cobb-Douglas production function elasticities for various inputs were derived indirectly from the profit function estimates (Table 3) using four-year (1967–68 to 1970–71) data, and are presented in Table 6. These estimates are obtained from identities which link the coefficients of the profit function and those of the production function. The main advantage of these indirect input elasticities over the ones obtained from direct estimates of the production function is their statistical consistency. Since β_1 appears in both the profit and labor demand equations, imposing the restriction that it be equal in both equations improves the efficiency of these estimates. Furthermore, since these estimates are derived from four-year data they should be quite reliable for predictive purposes.

All the estimates of output elasticities with respect to various inputs (including capital) have the expected signs and reasonable magnitudes. I seem to be fortunate in having data which yielded reasonable elasticity estimates for capital. L-Y obtained (because of the problem of measuring the capital input) negative elasticity for capital and, under constrained estimation with constant returns to scale, relatively large elasticity values for labor and land.

Secondly, whereas my findings agree with L-Y regarding equal relative price efficiency and equal absolute price efficiency of small and large farms, the findings regarding equal technical and thus equal overall economic efficiency differ. They find small farms relatively more efficient technically and thus more efficient economically, whereas my results indicate no differences in technical or economic efficiency of small and large farms.

TABLE 6—ESTIMATES OF THE INPUT ELASTICITIES OF THE COBB-DOUGLAS PRODUCTION FUNCTION DERIVED FROM THE PROFIT FUNCTION FOR NEW WHEAT 1967–68 to 1970–71,[a] PUNJAB, INDIA

		1 Restriction[b]	2 Restrictions[c]	3 Restrictions[d]
Labor	α_1	0.078	0.218	0.213
Land	α_2	0.636	0.547	0.522
Capital	α_3	0.349	0.280	0.265
$(\alpha_1+\alpha_2+\alpha_3)$		1.063	1.045	1.000

[a] Table 3, last column.
[b] See fn. a, Table 3.
[c] See fn. b, Table 3.
[d] See fn. c, Table 3.

A possible explanation for this discrepancy might be as follows:

Their findings pertain to the mid-1950's. Indian agriculture at that time could be characterized as traditional and in a state of equilibrium with available technology (see Schultz). Modern inputs like chemical fertilizers were conspicuous by their absence. Smaller farms which had more labor available per unit of land[12] perhaps used it for more intensive land improvement programs which resulted in superior technical efficiency compared to the larger farms. Also as emphasized by L-Y, under these circumstances the technical-managerial input becomes more intensive on smaller farms. Their finding of superior technical efficiency of smaller farms thus seems to be consistent with these observations.

Since the mid-1950's, however, Indian agriculture has undergone a great transformation, especially in Punjab. The level of land fertility which formerly depended on the level of labor input and could be higher on small labor-surplus farms no longer depends upon intensive labor input alone. The availability of fertilizers, other chemical inputs, and increased irrigation input reduces the fertility (productivity) differences of land on small and large farms. Thus a major source of greater technical efficiency of

[12] At this point a reference is made to the studies by Amartya Sen, the survey article by Jagdish Bhagwati and Sukhamay Chakravarty, and T. N. Srinivasan.

smaller farms during the mid-1950's seems to be less important during the late 1960's.

Another explanation can be advanced in the form of an hypothesis. There are two elements to this hypothesis. First, we may agree (in a somewhat qualified manner) with the findings of L-Y (1971, 1973) that in traditional agriculture or in an agriculture in a state of equilibrium, smaller labor-surplus farms have greater technical efficiency and thereby are more efficient economically. Second, we postulate that large farms have better access to research information because of relatively easier (often free) access to extension services. The period covered by the present study immediately followed the introduction of high-yielding varieties of wheat. Thus, it may well be that larger farms, because of their comparative advantage in research information, assimilated the new wheat technology more rapidly than smaller farms and this offset the technical superiority of smaller farms. This hypothesis can be verified only in the future.

D. *Elasticity Estimates*

Next I present a number of important elasticity estimates derived from the profit function estimates for four-year data (1967–68 to 1970–71), shown in the last column of Table 3. The labor demand elasticities with respect to wage rate w, land L, capital K, and price of output, respectively, are -1.271, 0.663, 0.337, and 1.271.

All these elasticity estimates have the expected signs. The price elasticity of demand for labor indicates that demand is quite responsive to wage levels. Positive responses for labor demand to increases of land, capital and output price have important implications for labor absorption in wheat farming.

The elasticities of output supply with respect to the normalized wage rate and output price are -0.271 and 0.271, respectively. The relatively inelastic output response with respect to wage rate along with an elastic response of demand for labor with respect to wage rate imply that exogenously enforced wage rates for agricultural labor above the market determined wage rates could result in substantial increase in unemployment of the agricultural labor force.[13] The magnitude of output response with respect to wheat price is important for any effort to use the output price variable as a policy instrument for inducing increased supply of wheat.

We also obtain the reduced form output elasticities with respect to land and capital, 0.663 and 0.337, respectively. These elasticities indicate the output response of the average farm with respect to exogenous increases in land and capital, respectively, holding the normalized wage rate and not the quantities of labor as constant. A given increase in the quantity of land (capital) shifts upward the marginal productivity curves of labor and other factors of production. As a result, more of these inputs are employed than before. Thus, holding wage rate constant (but not the quantities of labor), a 1 percent expansion in wheat land will result in 0.663 percent increase in wheat output and a 1 percent increase in capital will result in 0.337 percent increase in wheat output.

III. Summary and Conclusions

There are two substantive conclusions that follow from the analysis of our data. First, there seem to be limited possibilities for growth by improving allocative efficiency in moving toward production frontiers. This is the inference from tests indicating rational producer response to disturbances in the labor market generated by shifts in the labor demand function. On the other hand, technical changes such as the shift in the wheat production function on the order of about 45 percent, popularly known as the "green revolution," constitute the more important source for potential increases of output. Second, we find that tractor-operated wheat farms are no better off in terms of their economic performance than non-tractor-operated farms and that large farms are no better off than small farms. There are no differences in the technical and price efficiency parameters of these classes of farms. Policy for cur-

[13] For excellent discussions of the impending problem of labor force absorption in the context of the "green revolution" and farm mechanization, see Bruce Johnston and John Cownie, Billings and Singh, and C. H. Hanumantha Rao.

tailing farm size may be based only on social and political considerations. A qualification about this implication, however, is necessary because we have studied only the wheat crop out of the complete set of enterprises on Punjab farms. The picture may be different if we study the production relationship between aggregate output of all enterprises and the inputs used.

REFERENCES

J. S. Bhagwati and S. Chakravarty, "Contributions to Indian Economic Analysis: A Survey," *Amer. Econ. Rev.*, Sept. 1969, Supp., *59*, 1–73.

M. H. Billings and A. Singh, "The Effect of Technology on Farm Employment in India," *Develop. Digest*, Jan. 1971, *9*, 98–107.

———, "Conventional Energy as a Constraint to the Green Revolution 1964–84: The Punjab Case," mimeo. for USAID, New Delhi, Apr. 1969.

D. W. Hopper, "Allocation Efficiency in a Traditional Agriculture," *J. Farm Econ.*, Aug. 1965, *47*, 611–24.

B. F. Johnston and J. Cownie, "The Seed Fertilizer Revolution and Labor Force Absorption," *Amer. Econ. Rev.*, Sept. 1969, *59*, 569–82.

A. M. Khusro, "Returns to Scale in Indian Agriculture," *Indian J. Agr. Econ.*, Oct. 1964, *19*, 51–80.

L. J. Lau, "Some Applications of Profit Functions," memorandum no. 86A, 86B, Res. Center Econ. Growth, Stanford Univ., Nov. 1969.

——— and P. A. Yotopoulos, "A Test for Relative Economic Efficiency and Application to Indian Agriculture," *Amer. Econ. Rev.*, Mar. 1971, *61*, 94–109.

——— and ———, "Profit, Supply and Factor Demand Functions," *Amer. J. Agr. Econ.*, Feb. 1972, *54*, 11–18.

C. H. H. Rao, "Farm Mechanization in a Labor-Abundant Economy," *Econ. Polit. Weekly*, Feb. 1972, *7*, 393–400.

G. S. Sahota, "Efficiency of Resource Allocation in Indian Agriculture," *Amer. J. Agr. Econ.*, Aug. 1968, *50*, 584–605.

T. W. Schultz, *Transforming Traditional Agriculture*, New Haven 1964.

A. K. Sen, "Peasant and Dualism with or without Surplus Labor," *J. Polit. Econ.*, Oct. 1966, *74*, 425–50.

S. S. Sidhu, "Economics of Technical Change in Wheat Production in Punjab, India," unpublished doctoral dissertation, Univ. Minnesota 1972.

T. N. Srinivasan, "Farm Size and Productivity: Implications for Choice Under Uncertainty," disc. pap. no. 53, Planning Unit, Indian Statist. Inst., Jan. 1971.

P. A. Yotopoulos and L. J. Lau, "A Test for Relative Economic Efficiency: Some Further Results," *Amer. Econ. Rev.*, Mar. 1973, *63*, 214–23.

A. Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *J. Amer. Statist. Assoc.*, June 1962, *57*, 348–75.

Directorate of Economics and Statistics, "Studies in Economics of Farm Management, Ferozepur District, Punjab—Part II Tables," Ministry of Food and Agriculture, Government of India, 1967–68, 1968–69.

Economic Adviser, Government of Punjab, "Tractor Cultivation Scheme," unpublished data files, Chandigarh, Punjab, 1969–70.

Redistribution and the Pareto Criterion

By Harold M. Hochman and James D. Rodgers*

The view that income redistribution must be interpreted as a zero-sum game squares neither with what people say or do, privately or collectively, nor with the widely held view that amelioration of poverty is a "commendable aim of social policy." Thus, some six years ago, we turned to utility interdependence to examine, with conventional tools, both the positive and normative aspects of income transfers and showed through this device how some redistribution might make everyone better off and, therefore, be recommended by the Pareto criterion (see the authors (1969)). This analysis has, however, left E. J. Mishan (1972) decidedly unhappy. Indicting the Pareto optimal approach to redistribution on operational, conceptual, and philosophical grounds, he argues that this way of dealing with the normative implications of individuals' distributive preferences is "futile," "foredoomed to failure," and redundant to boot, since others have already given the subject of interdependent preferences "ample attention." It appears there is little more we could have done to offend Mishan's intellectual sensitivities. The purpose of this note is to clarify some of the issues raised by his critique and to expose the misconceptions on which much of it is based.

I. Mishan's Criticism and Our Arguments: Some Clarification

Mishan's critique contains three parts. The first asserts that " . . . attempts to derive distributional propositions from efficiency considerations are foredoomed to failure" (p. 972), because (a) such propositions have no operational value, and (b) they must, in any case, " . . . be referred ultimately to some *non-Pareto* criterion" (his italics, p. 974). The second argues that our approach would run into ethical difficulties, even if the first objection were put aside, because Pareto optimal distributional adjustments might aggravate rather than reduce measured income inequality. Finally, Mishan puts forward a " . . . stronger case for rejecting the interdependent utility approach to distribution—" claiming that, since welfare economics is " . . . founded on ethics, not utility" (p. 975), externalities arising from benevolence " . . . may not be agenda for society" (p. 976).

A device that may clarify our discussion of these criticisms, as well as the subject of our 1969 paper, is the "welfare frontier" or "utility-possibility locus," depicted for the two-person case in Figure 1.[1] The vertical and horizontal axes measure the utility levels of Mutt and Jeff, respectively; P_1P_2 is the welfare frontier. (Ignore for now the dashed curve $P_1'CP_2'$.) Our analysis deals with the situation in which, initially, the community finds itself at a point such as A on the upward-sloping portion. Here, as at all other points from P_1 to B, Mutt's preferences exhibit a Pareto-relevant externality sufficiently strong to cause the welfare frontier to slope upward. Point A, therefore, is not a Pareto optimal *point*. In the two-person case, this position is not likely to be stable. Mutt will simply make a transfer to Jeff (who, we assume, engages in no strategic

* Principal research associate, The Urban Institute, Washington, D.C., and visiting lecturer, Graduate School of Public Policy, University of California, Berkeley; and associate professor of economics, The Pennsylvania State University, respectively. We wish to thank, without implicating, Denis Aitken, E. C. Budd, Charles Cathcart, A. J. Cuyler, Robert Harris, George Peterson, and Richard Wagner for comments. We also acknowledge research support from the National Science Foundation.

[1] The welfare frontier is discussed lucidly by J. de V. Graaff (pp. 59–63, 71–74). Our Figure 1 has an upward-sloping segment at both "ends," reflecting a symmetry assumption: reversal of the roles of the two parties is assumed to have no effect on the results. We ignore the case where the frontier first slopes downward, then upward, then downward again. On this, see A. Mitchell Polinsky.

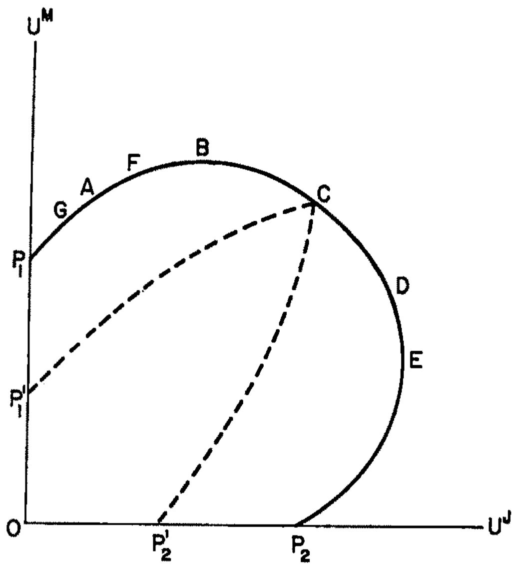


FIGURE 1

behavior), resulting in a move to B and making both parties better off. Such redistribution is, then, a Pareto optimal *move* to a Pareto optimal *point*. If there are many Mutts, on the other hand, the many-person analogue to point A may be a stable equilibrium, since private, cooperative action of the many Mutts may fail to occur because of the free-rider problem (or at least any private effort may not succeed in achieving sufficient redistribution to move the community to a position where the frontier no longer slopes upward). Hence, there arises a possible (though not necessarily conclusive) case for a governmental redistributive program, based not on ethical judgments about the way one person's welfare is to be compared to another's, but only on the Pareto criterion and using only its ethical judgments.[2]

Mishan, unfortunately, has interpreted us as asserting much more than this, apparently regarding our effort as purporting to "solve the distribution problem for society" (p. 974). Overlooking the limits of our normative claims, he seems to interpret us as saying that our approach can identify the *optimum optimorum*. But this is a misinterpretation. Our analysis deals only with Pareto improvements, such as the move from A to B in Figure 1, and not with the less tractable problem of choosing the best point from the set of "Pareto Optimal Distributions" (to note Mishan's mistitling of our paper, p. 972),[3] represented by the points on the welfare frontier from B to E. To assert that "bliss" or the point of maximum "social welfare" can be obtained by resort to the Pareto criterion is to assert that the Pareto criterion can rank Pareto optimal points, which it obviously cannot.[4] We do not, then try to "persuade [Mishan] that distribution [in all cases] need not be conceived as a separate aspect of a welfare criterion" (p. 973). Our remark about the collapse of equity into efficiency, to which Mishan refers, means only that in the situation we were analyzing, where the community finds itself on the segment P_1B in Figure 1, there is *no conflict* between the changes required to achieve Pareto optimality and principles of equity.[5]

II. Criticisms of Operational Value and Conceptual Validity

Mishan's criticism that a public program of Pareto optimal transfers would not be

[2] There are three of these: the welfare of each person counts; each person is the best judge of his own welfare; and an increase in one person's welfare with no reduction in another's is "good." We emphasize the ethical basis of the Pareto criterion lest anyone mistakenly conclude that Mishan's opening remarks concerning attempts to justify redistribution "scientifically" apply to us. We have nowhere asserted that *any* prescriptive statement flow from "some essential part of economics."

[3] We are not disturbed by this as semantic quibblers, but because it reflects Mishan's wrongheaded interpretation of our argument.

[4] It does not necessarily follow, however, that "Once the ethical basis [of allocative statements] is acknowledged . . . one cannot . . . determine the distribution of the product from a consideration of externalities" (p. 972). An example that contradicts this assertion, albeit one that employs an unrealistic, extreme assumption about benevolence, is provided in Figure 1. If benevolent externalities are symmetrical and sufficiently strong, the welfare frontier could appear as the dashed curve $P_1'CP_2'$, everywhere upward-sloping, save at a single point, C.

[5] At least no conflict would be registered by *any Paretian* social welfare function (see Graaff, pp. 8–10). As we note later, part of what Mishan says suggests that he thinks individuals either would not or should not desire to have social policy guided by a decision criterion that embraces the Paretian ethical judgments.

operational is a charge that could be levied with more or less equal force against a host of welfare propositions, particularly those involving external effects over large groups of people. However, the practical difficulties posed by the high costs of obtaining information about preferences and by possible nonuniformities in preferences at given income levels seem neither more nor less of a hindrance to political efforts to design a system of Pareto optimal transfers than they are to ascertaining the Pareto optimal tax-prices of the Lindahl solution where public goods and other nondistributional externalities are involved.[6]

Mishan's denial that our approach is conceptually valid is more troubling. He complains that since our analysis cannot enable us to identify the best point along *BD* (a problem that requires the "prior choice of a pretax income distribution"), its "more modest proposal of promoting Pareto-efficient redistributions . . . that start from a given distribution of real incomes" is useless (p. 974). It is true, of course, that our analysis deals only with the case in which property rights are predefined, and something is known about endowments, and not with the process of constitutional choice in which the property rights that underlie the pretransfer distribution are yet to be determined. However, recognition that we are advocating Pareto optimal *change* but have not solved the distribution problem hardly seems sufficient grounds for damning the recommendations—unless Mishan is really saying that piecemeal, sub-optimizing Pareto improvements may actually reduce social welfare. Unfortunately, this singularly unpalatable argument, which implies that any and all measures of social policy a society may contemplate must take on all of the most general and fundamental questions of political philosophy, seems to underlie much of Mishan's critique.

This brings us to a disturbing, though unstated, implication of this critique—if one accepts it. Our argument rests on refutable hypotheses about the selfsame preferences of individuals from which welfare economics, employing the Paretian ethical postulates, so comfortably derives its normative claims about resource allocation. To the extent that the same ethical premises underlie welfare propositions about allocation and distribution, denial of the distributional propositions calls into question the allocative statements, and in consequence, all normative content of neoclassical theory of resource allocation.

III. Redistribution That Increases Income Inequality

Mishan's argument that Pareto optimal transfers, given envy or nonuniformities in preferences, might increase income inequality merits brief comment. The issue with envy, within the limits of our frame of reference, is whether the transfers it suggests actually represent Pareto optimal moves. Where they do, the Pareto criterion, by itself, gives us no reason to deny them. For example, if Mutt and Jeff hate each other and would each feel better off were the goods and services available to each cut in half, and if this income reduction and the way it is achieved harms no one else, then bully for them and for the *increase* in social welfare that the income reduction produces. It is, after all, individual welfare levels and not incomes per se that are important. Again, if rich Mutt feels better off when poor Jeff is poorer (in money terms), and if Jeff (by some quirk) derives net benefit by making a transfer to Mutt, then let the transfer which benefits everyone take place. If, however, only Mutt were to benefit while Jeff were to suffer, the Pareto criterion would be silent.[7]

[6] The operational issue has been given a lucid and more thorough examination than that provided by Mishan. See George von Furstenberg and Dennis Mueller.

[7] Geoffrey Brennan discusses the implications of envy for Pareto optimal redistribution in a recent article and, interestingly enough, finds that in most cases the implied pattern of transfers would reduce inequality. To us, however, the likelihood that envy will be so important a factor in a person's preference as to make mutual income reduction a positive good seems low. Nor is envy, unlike benevolence, of much practical interest when the number of parties involved becomes large, so that opportunities for third-party harms multiply, or when discounted long-term implications are considered.

Provided that any negative external effects which might result from dissimilar tax treatment are ruled out, preference nonuniformities, like envy, would seem to pose no serious difficulties.

IV. Benevolence and Distributive Ethics

We come at last to Mishan's pièce de résistance—his rejection of the interdependent utility approach because " . . . welfare economics is founded on ethics, not utility." On this distinction, Mishan builds an argument that externalities of benevolence may not be "agenda for society" since benevolence may be inconsistent with distributive justice. Thus, "benevolence . . . would entail one set of transfers," while "the sense of justice would entail another" (p. 976).

Within the Paretian framework, these remarks could mean either of two things. Referring to Figure 1, they might mean 1) that benevolence calls for a transfer producing a move from A to B, but that justice requires a move from A to, say, D, which effectively modifies property rights in own income. Alternatively, they might be interpreted as meaning 2) that although benevolence entails the move from A to B, justice entails a smaller transfer from Mutt to Jeff, moving them from A to, say, F, or even a transfer from Jeff to Mutt, moving them from A to G.

In case 1), it can hardly be argued that distributive justice and benevolence are inconsistent, for attainment of maximum welfare (as registered by any *Paretian* social welfare function) requires that Pareto-relevant externalities of benevolence be eliminated.[8] Case 2) suggests the abandonment of the Pareto criterion, not for its failure to give strong enough policy recommendations, but for the advice it is capable of giving. To those of us who find the use of *Paretian* social welfare functions at all appealing, this implication is shattering. It implicitly asserts that no matter what *everybody* wants, *society* may prefer something else.[9]

There are, in addition, cogent economic reasons for questioning the alleged inconsistency between benevolence and justice. In economic transactions, ethics are manifested in rules, custom or code, which reduce uncertainties and constrain exchange. Benevolence, serving as an important stabilizing force in society, is one reason why people accede to such rules, even where it does not seem in their interest to do so when the rules are evaluated with reference to individual events.[10] A second, more conventional, reason is that ethical constraints reduce transactions costs. Under either interpretation, so long as society is beyond that "hypothetical state . . . in which no man knows what his income in life is to be," ethics appear to be a matter of utility, serving the Pareto criterion and reducing the harmful side effects of market transactions. In the post-constitutional setting, unless men are perverse, preferring less to more and suffering to pleasure, there seems to be no ethical basis for denying that Pareto optimal changes are just.

In arguing that benevolence and justice may be inconsistent, then, Mishan must be referring to the preconstitutional setting. Indeed, he appears to be claiming that individuals, even behind John Rawls' "veil of ignorance," might deny each other's rights to act on benevolence. Otherwise, why might "benevolence entail one set of transfers . . . [while] the sense of justice would entail another"? But for this to be true, either the

[8] That readers might regard possibility 1) as being relevant is what led us to remark that "one might personally feel that the amount of redistribution dictated by the Pareto criterion will not be 'enough.' We are not saying society should necessarily follow *only* the Pareto rule" (see our 1969 paper, p. 556).

[9] It should be noted that Mishan's examples on pp. 975–76 do little to build a convincing case for his position, since none of them demonstrates that the recommendations of the Pareto criterion lead to ethically deplorable results. The first does not even merit comment. The second, in which a shopkeeper bribes a gang of toughs, appears to be cogent support for his position, but it is actually no better. Legal stricture against bribery and extortion indicate that such transactions are not Pareto optimal from the standpoint of the community at large. As for Mishan's third example, see Section III above.

[10] To corroborate this argument, one need go no further than evidence of parental concern for children and collective agreement to rules circumscribing parental behavior.

state would have to be separate from or transcend its constituents, or the dominant human motivations, even in preconstitutional naiveté, would have to be envy and malevolence.[11] In the first case, it would be difficult to impart much meaning to the notion of a social contract; in the second, there would be nothing to assure its emergence.[12]

Fairness to Mishan, however, suggests a more charitable interpretation of his argument that externalities of benevolence may not be agenda for society. It could be argued, abandoning all neoclassical pretension, that an *emphasis* on the Pareto criterion and the ends and means of Pareto optimal redistribution may divert society from dealing with the issue of *optimal distribution* and from giving due consideration to fundamental social ills. In terms of Figure 1, energy spent on moving from A to B may reduce the chances of attaining "bliss" at, say, D. This radical argument, to be sure, carries a dim ring of truth, though it casts a dark shadow on Paretian welfare economics. To those of us who wonder how criteria of justice are to be set, and by whom, this argument is deeply troubling.

V. Conclusion

To us, Mishan's discomfort with utility interdependence as a rationale for income transfers is very much bound up with his earlier distinction between "psychic" externalities, like the interdependencies we discuss, and "tangible" externalities, which impinge directly on consumption or production. Mishan (1969) dismisses psychic externalities as extraneous to economics.[13] While this may be convenient, it is useful only if one's purpose is to retain a clean taxonomic division among the social sciences, confining economic reasoning to its traditional subject matter. To us, it seems more reasonable to interpret economic concepts as logical tools, useful in any discussion of constrained choice, whether the objects of choice are tangible or not. To the economist, psychic externalities should be no less relevant than tangible externalities, so long as he believes that their examination serves to enrich positive knowledge by suggesting nontrivial hypotheses and producing useful insights into behavior. Gary Becker, the authors (1973), Rodgers, and Robert Schwartz, among others, discuss such hypotheses.

Shorn of its misinterpretations, Mishan's argument, which sets out to discredit the concept of Pareto optimal redistribution, succeeds only in reminding us, unnecessarily, that neither economics nor any other science can resolve questions of values and that there is some range of income distributions the Pareto criterion cannot rank. Still, while Mishan's case is innocuous, the manner in which he has made it is pernicious. He has presented an articulate defense of the methodological status quo which founders, substantively, on a mystical Maginot line between psychic and tangible externalities, and on his denial that efforts to adapt economic theory to reflect such positive phenomena as utility interdependencies are worthwhile. In our experience and observation, utility interdependencies are real. This is our fundamental reason for arguing that economic analysis should take them into account. Would not disinterested persons, acting behind the "veil of ignorance," agree?

REFERENCES

G. S. Becker, "A Theory of Social Interactions," unpublished paper, Univ. Chicago, Sept. 1969.

[11] This is, of course, implausible, for in the "hypothetical state prior to . . . existing worldly interests" no one would yet have anything for others to envy.

[12] See Rawls for a far-reaching effort to develop a "social contract" theory which defines the preconditions and rules of justice. For Mishan's benefit it should be noted that the rule changes required to bring about a just society are contingent on benevolence in Rawls' framework. Only in the "state of nature" can nonbenevolent individuals be counted upon to submit to the contractarian exercise that justice requires.

[13] In his 1969 article, Mishan asserts, without argument or substantiation, that " . . . they [psychic externalities] play no part in economic policy and are ignored in all cost-benefit studies" (p. 330). But whether they *do* play such a part is a question of fact and whether they *should* play a part, if they exist, is a matter of judgment. That they *do* and *should* is the thrust of the literature with which Mishan is disturbed.

G. Brennan, "Pareto Desirable Redistribution: The Case of Malice and Envy," *J. Publ. Econ.*, Apr. 1973, *2*, 173–83.

J. de V. Graaff, *Theoretical Welfare Economics*, New York 1957.

H. M. Hochman and J. D. Rodgers, "Pareto Optimal Redistribution," *Amer. Econ. Rev.*, Sept. 1969, *59*, 542–57.

——— and ———, "Utility Interdependence and Income Transfers Through Charity," in K. E. Boulding et al., eds., *Transfers in an Urbanized Economy: Theories and Effects of the Grants Economy*, Belmont 1973, 63–77.

E. J. Mishan, "Joint Products, Collective Goods, and External Effects," *J. Polit. Econ.*, May/June 1969, *77*, 329–45.

———, "The Futility of Pareto-Efficient Distributions," *Amer. Econ. Rev.*, Dec. 1972, *62*, 971–76.

A. M. Polinsky, "Shortsightedness and Nonmarginal Pareto Optimal Redistribution," *Amer. Econ. Rev.*, Dec. 1971, *61*, 972–79.

J. Rawls, *A Theory of Justice*, Cambridge 1972.

J. D. Rodgers, "Explaining Income Redistribution," in H. M. Hochman and G. E. Peterson, eds., *Redistribution Through Public Choice*, New York 1974.

R. A. Schwartz, "Personal Philanthropic Contributions," *J. Polit. Econ.*, Dec. 1970, *78*, 1264–91.

G. M. von Furstenberg and D. C. Mueller, "The Pareto Optimal Approach to Income Redistribution: A Fiscal Application," *Amer. Econ. Rev.*, Sept. 1971, *61*, 628–37.

A Note on Welfare Surpluses and Gains From Trade in General Equilibrium

By James E. Anderson*

A recent article in this *Review* by E. J. Mishan (1968) and a subsequent exchange between Mishan (1971) and Melvin Krauss and David Winch have clarified to some extent ambiguities in Harry Johnson's famous attempt to resuscitate consumer's and producer's surplus measures of gains from trade. Nevertheless, a crucial issue remains obscure. Mishan claims that Johnson's identification of the welfare effects of a tariff as an excess of consumer's surplus loss over producer's surplus gain is arbitrary with respect to *concept*, hence improper. All that can be identified is a net loss, which is measured purely in terms of consumer's surplus. In the particular case of two goods, Mishan's objection reduces to denying any significance to the area in back of the general equilibrium supply function. Krauss and Winch side with Johnson, but unconvincingly. The main problem is that, though all writers deal in general equilibrium terms, their arguments are shackled by the use of the ambiguous surplus concepts. The purpose of this note is to present the gains from trade measurement problem in a consistent general equilibrium framework. The measure of the gains from trade attained is very similar to Johnson's: the net gain (measured correctly by all) is a *difference* between a change in income from production and a change in income from the income effect of price changes in consumption. Both sources of gain (or loss) are evaluated through their impact on social welfare, but the identification of the components of the net gain (loss) by source is natural in terms of the decision problem facing the general equilibrium economy. In the two-good case analyzed in the exchange, the area in back of the general equilibrium supply function measures the change in income from production and thus has a very useful interpretation. To identify the production-income change with changes in producer's surplus and the income effect on consumption with changes in consumer's surplus is, however, anachronistic and misleading and in any case unnecessary.

* Associate professor, Boston College. I am indebted to John G. Riley for helpful comments.

I

The essence of the approach to measuring gains from trade is to see that the change in domestic and international prices has two distinct effects: 1) at constant money income there is a change in real income (Hicks' compensating variation), and 2) there is a change in money income. The latter translates into a change in real income, of course, but it is very good sense to consider the two effects separately. Proceeding more formally, if government revenue is assumed to be optimally redistributed in lump sum fashion according to a Bergson welfare function, the general equilibrium of the competitive economy may be characterized as the simultaneous solution of two maximization problems: 1) for given producer and international prices (not necessarily the same), maximize money national income (the value of production and trade), and 2) for given consumer prices and money national income, maximize social utility. Changes in any of the three sets of prices will in general disturb the solution to both problems, resulting in a new level of utility. We define the change in real national income as the change in money income necessary to maintain the old level of utility holding consumer prices constant at their new values. It is a difference between two conceptually distinct changes—the change in money income from production and trade, and the change in money income which compensates for the change in consumer prices.

The specific case originally analyzed by Johnson and repeated here is that where changes are induced by the imposition of a tariff. Consumer and producer prices are

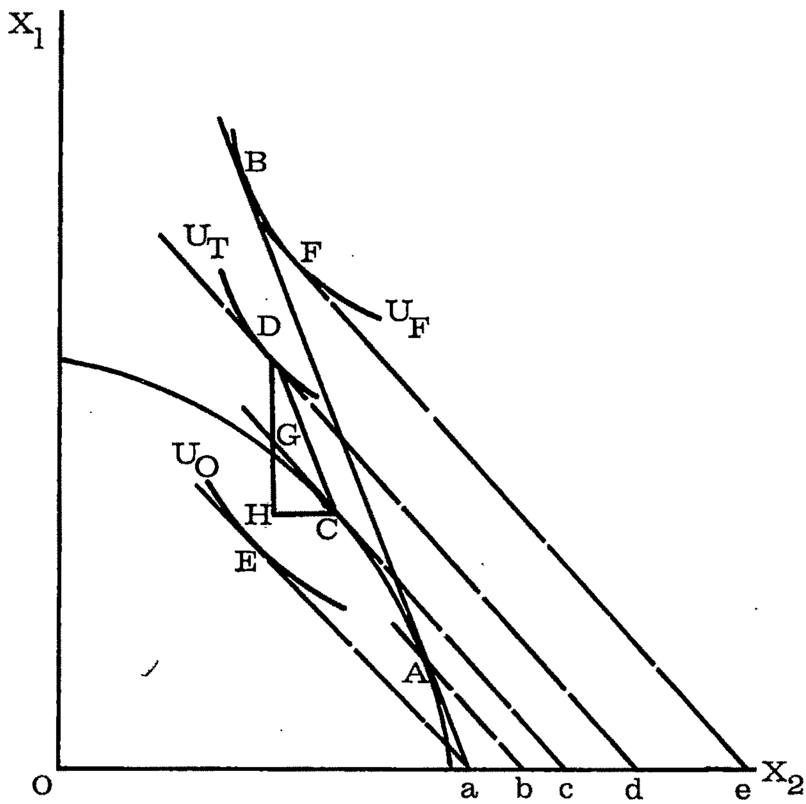


FIGURE 1

equal, with the trade tax creating a divergence between them and international prices. The trade tax results in government tariff revenue which is part of the income from trade and production. Before proceeding with a general mathematical treatment, it will be illuminating to explore Johnson's two-good case graphically (see Figure 1). The essentials are the same as those in Mishan's analysis of free trade vs. autarky. All dashed lines are parallel. Tangencies obtain at A, B, C, D, E, and F. The points C and A are on the efficient transformation surface. For convenience, international prices are assumed fixed, since nothing essential is gained by letting them vary. Free trade allows production at A and consumption at B with X_2 exported for X_1 at the fixed international terms of trade; U_F and U_T are the free trade and protected levels of (ordinal) social utility. A tax on the import of X_1 moves the production point to C and the consumption point to D. The net loss of income in terms of X_2 (evaluated at protected prices) is de. The consumption component of net loss is ae: it is the income change (in terms of X_2) which compensates for the price change (an increase in the consumer's price of good 1). Without an income change, U_O would be the new level of utility. This loss is partially balanced by a rise in money income at protected prices in terms of X_2. At the free trade production point with protected prices, X_1 is worth more in terms of X_2; the resultant change in income is ab. Output of X_1 expands causing a further increase in income, bc. Finally, trade has a tax on it, resulting in government revenue of cd.[1] The net of the production and trade

[1] Formal proof is simple, but omitted since examining Figure 1 is convincing. Protected equilibrium involves offering HC of X_2 for HG of X_1 domestically, while the international economy pays HD of X_1 for HG of X_2. The component GD of X_1 is the tariff revenue, worth cd in terms of X_2.

gain and consumption loss is thus de.

It is readily seen that this breakdown is not "arbitrary," since it involves only a separation of the effect of price changes into their impact on income-generating and income-consuming activities of the economy. In higher dimensions the case for this approach is still more persuasive, because the generalization of Mishan's method, while correctly measuring the net gain, obscures the working of the economy. We therefore now consider a general mathematical model.

On the production side we note that production income π is $\sum_i P_i X_i$. Also, with efficient production, $\sum_i P_i dX_i = 0$. The $X_i = X_i(P_1, \ldots, P_n)$ are the general equilibrium supply functions. Expanding π in a Taylor's series about the free trade point and neglecting higher order terms:

$$d\pi = \sum_{i=1}^{n} \frac{\partial \pi}{\partial P_i} dP_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 \pi}{\partial P_i \partial P_j} dP_i dP_j$$

Efficient production implies:

$$\frac{\partial \pi}{\partial P_i} = X_i + \sum_{j=1}^{n} P_j \frac{\partial X_j}{\partial P_i} = X_i$$

and thus

$$\frac{\partial^2 \pi}{\partial P_i \partial P_j} = \frac{\partial X_i}{\partial P_j}$$

$$d\pi = \sum_{i=1}^{n} X_i dP_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial X_i}{\partial P_j} dP_i dP_j$$

Finally, note that $I = \pi + G$, where G is net government revenue, and hence $dI = d\pi + dG$.

On the consumption side we follow J. R. Hicks, pp. 330–32, and denote consumer expenditure as M. Social utility is $U(C_1, \ldots, C_n)$, $U_i = \partial U / \partial C_i = \mu P_i$, $M = \sum_{i=1}^{n} P_i C_i$ and expanding M about the free trade point

$$dM = \sum_i \frac{\partial M}{\partial P_i} dP_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 M}{\partial P_i \partial P_j} dP_i dP_j$$

With utility constant,[2] for each good i,

$$dU = \sum_{j=1}^{n} \frac{\partial U}{\partial C_j} \frac{\partial C_j}{\partial P_i} dP_i = \sum_{j=1}^{n} \mu P_j \frac{\partial C_j}{\partial P_i} dP_i = 0$$

Therefore we note

$$\frac{\partial M}{\partial P_i} = C_i + \sum_{j=1}^{n} \frac{\partial C_j}{\partial P_i} P_j = C_i$$

$$\frac{\partial^2 M}{\partial P_i \partial P_j} = \frac{\partial C_i}{\partial P_i}$$

$$\text{and } dM = \sum_{i=1}^{n} C_i dP_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial C_i}{\partial P_j} dP_j dP_i$$

The change in real income, dR, is a net of dI and dM.

$$dR = dI - dM = \sum_i (X_i - C_i) dP_i + dG + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial X_i}{\partial P_j} dP_j dP_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial C_i}{\partial P_j} dP_j dP_i$$

With all goods traded at fixed prices in initial free trade, the dP_i represent trade taxes or subsidies. Government revenue will depend on actual consumption in the protected situation, which is not measured by movements on the free trade utility surface (i.e., by movements of compensated demand functions). If we simply disregard this difference in protected consumption amounts, $dG = -[\sum_i (X_i - C_i) dP_i + \sum_i (dX_i - dC_i) dP_i]$. The change in real income dR is then simply the negative of the last two terms.[3] Note that

[2] We hold utility constant at the free trade level. This is the simplest diagrammatic case, but we could equally well pick the protected level and obtain the effect of free trade prices, or use a combination of the two methods, obtaining a different measure in each case. The measure from the protected level is, of course, Hicks' equivalent variation. With normality, the measure of compensating variation from the free trade level will be the greatest, and from the protected level the least of the possible measures.

[3] A justification for this arbitrary procedure lies in our

this formula differs from Johnson's only by the presence of cross effects in supply.[4]

Generalization of Mishan's approach, focussed only on net gain, evidently utilizes compensated *excess* demand functions. We simply measure Hicks' compensating variation for the trade indifference function maintaining the free trade level of trade indifference. Government revenue will be mismeasured as above and the value of dR will be identical. The disadvantage of this procedure, apart from failing to make a very useful distinction, is that the supply and demand slopes are mixed in together. For both reasons, the working of the economy is obscured. Analysis of commodity taxes, where consumer and producer prices are not equal, of course requires the procedure above.

Specializing back down to the two-good case, we make good 2 the numéraire ($dP_2=0$) and obtain the net loss (minus the third and fourth terms of dR) *either* as *de* on Figure 1 *or* as area *abc* plus area *def* on Figure 2.[5] In the latter, it is the net of the increase in production income $P_1^*(1+t)caP_1^*$, the increase in government revenue *bcde*, and the compensating variation $P_1^*\ (1+t)dfP_1^*$, all measured in terms of X_2. This is essentially Krauss and Winch's procedure, but their analysis is obscure because they interpret the area in back of X_1 as an increase in factor rents. It is not actually wrong, since in general equilibrium the production income must

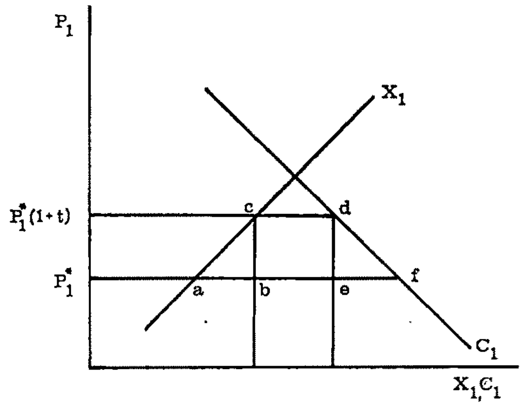


FIGURE 2

equal the total factor payments, but it draws one's attention from the main line of argument, particularly since they fail to make the appropriate division between income producing and income consuming activities. Mishan very properly says: "Neither the welfare of factors nor of firms can be distinguished as isolated from the original indifference transformation construction, since they are not there to begin with" (p. 204). Nevertheless, simply reasoning from the original construction (and its upper dimension analogue), the area in back of the general equilibrium supply functions has a satisfying and useful interpretation at the proper level of analysis, that of the economy.

The exchange between Mishan and Krauss and Winch also has a rather confusing discussion of the choice of numéraire. Mishan notes that, depending on the numéraire, a change in prices registers either as an excess of producers surplus loss over consumer's surplus gain or as an excess of consumers surplus loss over producer's surplus gain. For him, this apparently indicates the arbitrary nature of the conceptual base of the breakdown. Clearly, as Krauss and Winch say, the choice should not matter, but their analysis is obscure. In the two-good case the issues are illustrated as follows. In terms of our Figure 1 we can regard the move from U_F to U_T as caused by a *fall* in the price of good 2 rather than a rise in the price of good 1. With initial free trade it is convenient to make

previous recognition that the compensation measure is itself arbitrary. If we had measured the compensating variation at the protected level of utility (i.e., performed the experiment of going from protection to free trade), the protected consumption amount would be exact, and government revenue would cancel out as above. With normality, this would give the lowest possible measure of dR. The above procedure gives a measure between the highest and lowest measures.

[4] This is a serious slip in Johnson, since zero cross effects everywhere in supply is surely a pathological case.

[5] Simple geometry establishes that area abc = the third term of dR and area def = the fourth term of dR. These are, respectively, Johnson's production and consumption components of the net loss. Even Mishan admits the sense of this distinction. Note again that this diagram is inexact because *bcde* does not actually measure the tariff revenue, it is simply an approximation for convenience.

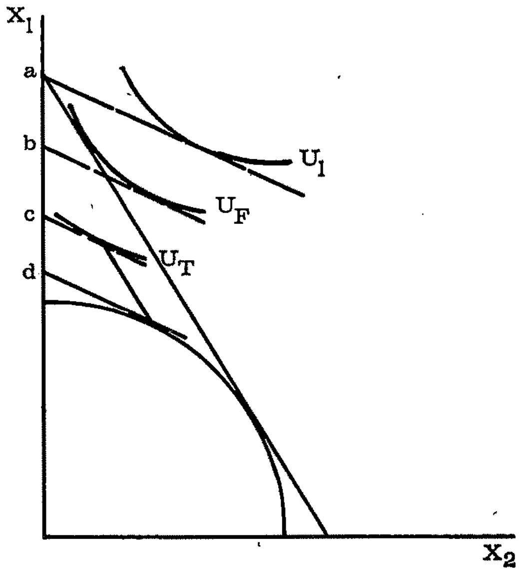


FIGURE 3

this due to an export tax of the same rate as the import tariff (by Lerner's symmetry theorem it has the same effect). At constant money income (in terms of X_1) there is a gain in real income to consumers. In Figure 3 this is measured by *ab*. There is a loss of money income from production of *ad*, and a gain of government revenue of *cd*, with a net loss of *bc*.

Despite Mishan's claims, there is nothing destructive of Johnson's breakdown in the fact that with good 1 the numéraire, we have a consumption loss and production gain while with good 2 the numéraire, we have a consumption gain and production loss. In any particular case of application we always know the structure of absolute price changes to insert in dR and hence which way to "properly" regard the components of the net gain or loss. Nevertheless, the homogeneity of degree zero of the general equilibrium supply and demand functions permits arbitrary choice of numéraire. We may set any one of the price differentials equal to zero and change all other price differentials so as to preserve the structure of relative price changes. Depending on the numéraire, we can change the sign of $d\pi$ and dM, with no significance.[6]

II

Measurement of the gains from trade using either Mishan's method or Johnson's as expanded here is far from our reach, so that work like that of Gerald Lage or H. David Evans is probably as good as we can expect. Johnson's approach has a misleading appearance of applicability, but retains at least a theoretical appeal. We are indebted to Mishan for pointing out his slip in calling the change in production income producer's surplus; otherwise it remains perfectly useful.

[6] Krauss and Winch appear to be troubled by the problem of possible double counting. One must of course measure the net gain either as *bc* on Figure 3 or *de* on Figure 1. But the formula for dR is impervious to double counting due to the homogeneity of degree zero of supply and demand functions, so it is difficult to see how anyone could be led astray using it.

REFERENCES

H. D. Evans, "Effects of Protection in a General Equilibrium Framework," *Rev. Econ. Statist.*, May 1971, *53*, 147–56.

J. R. Hicks, *Value and Capital*, 2d ed., Oxford 1946.

H. G. Johnson, "The Cost of Protection and the Scientific Tariff," *J. Polit. Econ.*, Aug. 1960, *68*, 327–45.

M. Krauss and D. Winch, "Mishan on the Gains from Trade: Comment," *Amer. Econ. Rev.*, Mar. 1971, *61*, 199–200.

G. Lage, "A Linear Programming Analysis of Protection," *Western Econ. J.*, June 1970, *8*, 167–85.

E. J. Mishan, "What is Producer's Surplus?" *Amer. Econ. Rev.*, Dec. 1968, *58*, 1269–82.

———, "Mishan on the Gains from Trade: Reply," *Amer. Econ. Rev.*, Mar. 1971, *61*, 202–07.

The Coase Proposition, Information Constraints, and Long-Run Equilibrium

By WILLIAM SCHULZE AND RALPH C. D'ARGE*

Ronald Coase in his now famous paper on social costs asserted that, given the initial "... delimitation of rights ... the ultimate result (which maximizes the value of production) is independent of ... [the choice of a liability or nonliability rule] ... if the pricing system is assured to work without cost" (p. 8). Guido Calabresi recently restated the Coase proposition more succinctly as follows: "The same allocation of resources will come about regardless of which of two joint cost users is initially charged with the cost, in other words, regardless of liability rules" (1968, p. 67). Of course, Calabresi also presumed zero transactions costs. In addition, Calabresi reiterated the Coase conclusion that liability rules do not influence the efficient allocation of resources in either the short or long run.[1]

A major logical problem with the Coase proposition is in introducing a perturbation into an otherwise perfect system and then asserting that regardless of legal rules, the system will regain its perfection. Coase, by introducing a perturbation, negates one of the assumptions necessary for asserting that the allocation of resources will be efficient regardless of liability rules in both the short and long run, namely perfect information.[2] A perturbation occurs which is not known in advance, i.e., information is lacking; but there is apparently complete information at zero cost for all future actual and potential exchanges.[3]

Much of the confusion on externalities and legal rules, we submit, has arisen because writers in the field have not made clear their assumptions on transactions costs.

Applying a competitive model of two industries with an externality and allowing for entry and exit we find that a liability rule will lead to a misallocation of resources in the long run, given the acceptance of a particular set of assumptions on transactions costs which are commonly accepted in the economics literature. We also find that a nonliability rule leads to misallocation. A Pigovian tax, alternatively, leads to a local optimum in resources allocation provided the system is stable and receptors are not compensated.

I

In this section we undertake a comparison of the Coase proposition and Pigovian taxes in the long run where particular positive transactions costs are introduced. That effort is concentrated on external diseconomies of

* Assistant professor of economics, University of New Mexico, and professor of economics, University of California, Riverside, respectively.

[1] See Guido Calabresi (1968, p. 68). Coase states "... the long-run equilibrium position ... is the same whether or not the cattle-raiser is held responsible for the crop damage brought about by his cattle" (p. 8).

[2] In the short run, bargaining solutions with or without a liability rule are Pareto efficient. However, a number of writers including Calabresi, and indirectly David Bramhall and Edwin Mills, have raised questions as to the conclusion by Coase that long-run allocation of resources will be the same regardless of liability rules. As noted earlier, Calabresi recently reversed his opinion and concluded that long-run as well as short-run resource allocation would be the same. See Calabresi (1968), and Bramhall and Mills.

[3] The question arises as to the nature and meaning of perturbations. There appear to be two categories of causes for such perturbations: purely random events much like accidents; and forms of designed coercion. For example, if farmer A's cows enter farmer B's corn field, Coase never analyzes why the cows entered. Did farmer A inadvertently allow entry or was it by design, i.e., A herded them close to the field so the cows would be attracted by the corn? Most of the literature on externalities takes it as given that the externality (or perturbation) results from unexpected or inadvertent behavior of one party. But, again, inadvertent behavior is suggestive that there *are* gaps in information and transactions costs are not zero. And coercion by design, other imprefections aside, cannot occur unless at least one party has less than perfect information.

production with firm entry where there are some implicit assumptions on transactions costs normally presumed in the literature on perfect competition. One such assumption is that firms and consumers are price takers. Also, it is taken as given that atomistic firms enter or exit a particular industry on the basis of expected future profits which equal, by assumption, current profits. Thus there is assumed to be an informational barrier (very high transactions costs) associated with knowledge of future profits. Further, it is assumed firms in both emitter and receptor industries can costlessly negotiate, but that third parties are barred from such negotiation. Finally, it is assumed that there is a government agency in the Pigovian tax case that can enforce tax payments and distribute revenues costlessly. These assumptions, while semi-inconsistent as regards transactions costs (i.e., some transactions costs are presumed to be extremely high and others virtually zero), underlie most studies of competitive behavior of firms.

It has been widely demonstrated that the Coase proposition is essentially valid, given the assumption of zero transactions costs after the appearance of an externality, in the short run. However, the "wealth effect" coupled with transactions costs which Coase, Calabresi, and others ignore is particularly significant when one considers competitive long-run equilibrium in production. Here, since the structure of industry is not fixed but is subject to a kind of Malthusian law of entry, not only are resources allocated most efficiently among firms, but the most efficient combination of firms is also assured in the long run under perfect competition with no externalities. The question to be resolved is whether or not negotiations between producers exploiting the gains from trade made available by an externality in production will result in an efficient solution not just in the short run but in the long run as well. Since profitability of firms determines entry (presumed to be costless), and since the allocation of property rights with respect to an externality will affect relative profitability of emitters and receptors, it is clear that the structure of industry will not be invariant to the configuration of property rights.[4] Further, free entry may be inconsistent with stability of a negotiated solution arising under a nonliability rule because new entrants can act as free riders, a possibility excluded by a liability rule or a Pigovian tax.

Consider a two-industry partial equilibrium model with a diffuse externality such that the output of industry 2 adversely affects the production of every firm in industry 1 in a like manner.[5] An example might be the release of air pollutants with instantaneous horizontal mixing by industry in an airshed which also contains agricultural producers. The question would then be to determine how many industrial firms and how many farmers should locate in the airshed. Where there are n_1 identical firms, each producing output y_1 in industry 1, and n_2 identical firms, each producing output y_2 in industry 2, we can write inverse demand functions (price P_i as a function of industry output n_iy_i) for industry 1 as $P_1(n_1y_1)$ where $P_1'(n_1y_1)<0$ and for industry 2 as $P_2(n_2y_2)$ where $P_2'(n_2y_2)<0$.[6] The total production cost for each firm in industry 1 is given as $C_1(y_1)+D_1(n_2y_2)$ where $C_1(y_1)$ is the direct cost of production for each firm in industry 1 and $D_1(n_2y_2)$ is the damage incurred by each firm in industry 1 as the result of the emissions of industry 2. Note that we are presuming damages are separable. However, some of our results concerning the optimality of various policy measures are not dependent on this assumption. Total cost for each firm in industry 2 is $C_2(y_2)$. We assume C_1', C_2', C_1'', $C_2''>0$ and D_1', $D_1''>0$ in the relevant regions of production for each industry.

The conditions for a social optimum in a partial equilibrium framework are generated by maximizing net benefits (NB) which can be defined as the difference between willingness to pay for the output of both industries

[4] This has been pointed out by Bramhall and Mills.

[5] We have also considered a two-industry general equilibrium model which gives essentially identical results. The approach used here facilitates the use of graphs and simplifies dynamic analysis.

[6] Primes denote differentiation.

and the total cost of production in both industries. Thus we seek a maximum of

$$(1)\quad NB=\int_0^{n_1y_1}P_1(s_1)ds_1+\int_0^{n_2y_2}P_2(s_2)ds_2 - [n_1(C_1(y_1)+D_1(n_2y_2))+n_2C_2(y_2)],$$
$$y_1,\ y_2,\ n_1,\ n_2\geq 0$$

where s_1 and s_2 are dummy variables of integration in the demand functions for total output of each industry. Assuming an interior solution, the first-order conditions are:

$$(2)\quad \partial NB/\partial y_1 = n_1(P_1 - C_1') = 0$$

$$(3)\quad \partial NB/\partial y_2 = n_2(P_2 - C_2' - n_1D_1') = 0$$

$$(4)\quad \partial NB/\partial n_1 = P_1y_1 - (C_1 + D_1) = 0 = \pi_1^*$$

$$(5)\quad \partial NB/\partial n_2 = P_2y_2 - (C_2 + n_1D_1'y_2) = 0 = \pi_2^*$$

The interpretation of (2) and (3) is quite straightforward and implies, where n_1, $n_2>0$, that for each firm in industry 1 price should be equal to marginal cost (C_1'), and that for each firm in industry 2 price should equal marginal cost (C_2') plus marginal damages to industry 1 (n_1D_1'). These are the usual short-run conditions with a unidirectional externality between firms. The conditions for a long-run optimum, (4) and (5), are more interesting since they should correspond to the definition of zero profits for firms in industries 1 and 2, respectively (assuming firms enter until profits are zero). Equation (4) implies that $\pi_1^*=0$ is the optimum level of profits where receptor firms bear the full cost of the externality D_1 at the optimum. This result suggests that compensation for damages will distort long-run equilibrium in the receptor industry. Equation (5) implies that $\pi_2^*=0$ is the optimum level of profits where the emitters must bear an additional cost of n_1D_1' per unit of output y_2 produced. This can be interpreted as an optimum long-run Pigovian tax equal to marginal damages on the output of the emitter industry. We note then that the optimal policy barring renegotiation after taxation is to do nothing with respect to the receptor firms, allowing them to bear the cost of the externality after the optimal tax on output has been applied to firms in industry 2.[7] This will assure the optimum number of firms in each industry. However, to prevent further distortion, the tax revenue must be distributed in the form of lump sums to consumers. Tax revenue *cannot* be used to compensate receptors unless it is unrelated to their view of damages.

The relationship between the optimal Pigovian tax case (denoted *), the unadjusted externality case (denoted E), and the liability case (denoted L) can be best demonstrated with the aid of Figures 1 (a firm in industry 1) and 2 (a firm in industry 2). In Figure 1, the optimal long-run equilibrium point for receptor firms is * at the lowest point of the average total cost curve *including* optimal damages suffered (AC_1^*). This point is defined by the intersection of the marginal cost curve (C_1') with the adjusted average total cost curve (AC_1^*). As damages (D_1) increase with increasing output of industry 2, the average total cost curve of firms in industry 1 including damages shifts upward. We presume that where the optimum tax is applied to firms in industry 2 and free entry exists for both industries, there will be a convergence to D_1^* the optimal long-run level of damages, and optimal price P_1^* and quantity y_1^* will result from the long-run equilibrium point (*). The dynamic stability of the Pigovian tax system will be examined later.

The optimum equilibrium point in Figure 2 for firms in industry 2 is also denoted *. This can be defined by the intersection of the average total cost curve including the tax (AC_2^*) with marginal social cost ($C_2'+n_1D_1'$). Note that this point corresponds to the zero profit condition for firms in industry 2 where AC_2^E is the adjusted average cost and the area $P_2^EE * P_2^*$ is the optimal long-run tax collected from each firm in the emitter industry. This implies that, if through some mechanism not involving a tax or levy, the

[7] See Allen Kneese and Blair Bower, p. 100, and Ralph Turvey.

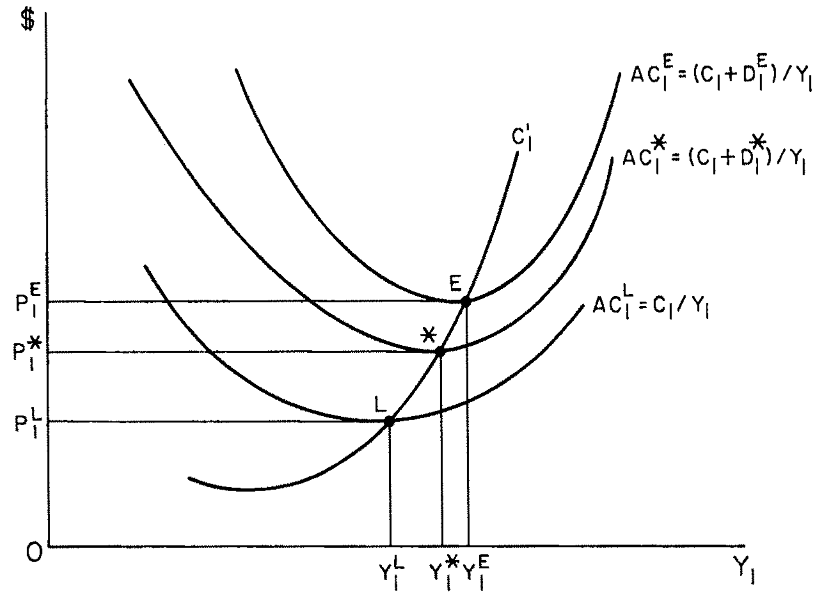


FIGURE 1

two industries reach the optimum point * in Figures 1 and 2, positive profits equal to the area $P_2^E E * P_2^*$ times n_2^* will be obtained. Since positive profits will induce more firms to enter, the optimum points * cannot be a stable equilibrium under free entry. Thus the Pigovian tax serves to remove these destabilizing profits which explains the asymmetry (taxes without compensation) which so bothered Coase.[8]

The uncompensated externality case results in a long-run equilibrium at point E in Figure 1 for firms in industry 1. Here, since damages received (D_1^E) will be greater than optimal (D_1^*), the average cost curve (AC_1^E) will lie above AC_1^*. Since the externality is separable, the marginal cost function (C_1') does not shift so entry or exit occurs until profits are zero resulting in a long-run price of P_1^E and output per firm of y_1^E. This implies, where the demand function for industry 1 as a whole is downward sloping ($P_1' < 0$), that total industry output and number of firms will be less than optimal in the unadjusted externality case for receptors, since externality price and output for each firm are both greater than optimal price and output per firm, respectively. This occurs because at the higher (nonoptimal) price, industry demand is less. In Figure 2 the long-run externality equilibrium point E for firms in the emitter industry occurs where average private cost (AC_2^E) equals private marginal cost (C_2') resulting in a price (P_2^E) lower than the optimum price (P_2^*). However, firm size is still optimal since output (y_2^E) in this case is identical to the case under taxation (y_2^*). The intuitive explanation of this result, which is not dependent on separability, is simply that in spite of the externality, social product is still maximized by producing each unit of y_2 as cheaply as possible. This implies that in the unadjusted externality case, there will be too much total output from the emitter

[8] Coase in fact proposed a double tax system, one tax on emitters equal to *total damages*, and another to regulate entry, p. 41. The optimum tax is of course determined by *marginal damages*, and this is the source of his confusion. More recently, William Baumol has attempted to demonstrate that a single Pigovian tax is needed to achieve a local optimum in terms of location and efficiency. Baumol's finding reinforces the conclusions identified here. F. Trenery Dolbear, Jr., p. 99, criticizes marginal taxes where there is an externality in consumption because emitters pay more taxes than damages induced. In production with free entry or in consumption where a location decision is involved, this *must* be the case to achieve the optimum externality.

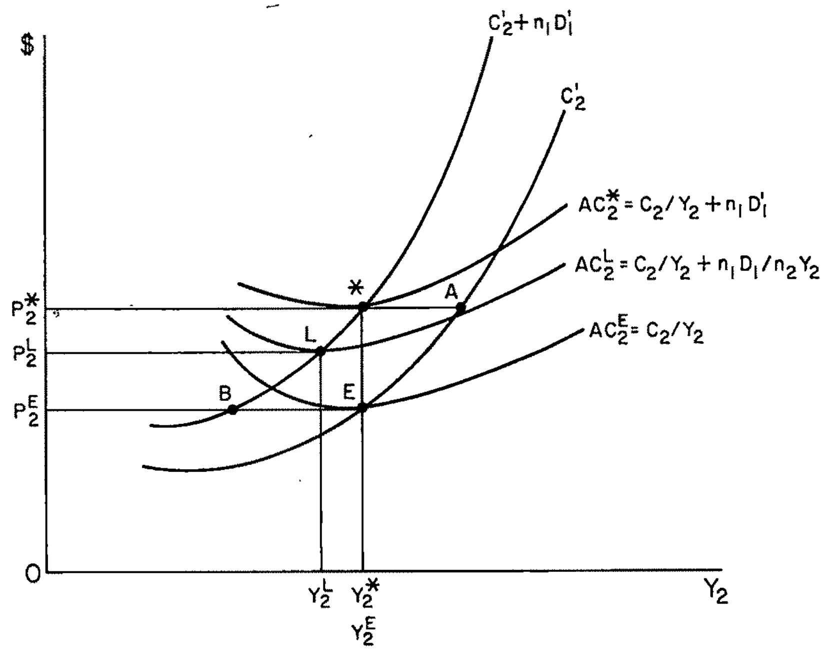


FIGURE 2

industry and too many firms, because the demand curve for the entire industry is assumed to be downward sloping ($P_2' < 0$) even though each firm perceives demand as infinitely elastic. Thus, in the unadjusted externality case, there is an underallocation of resources to the receptor industry and an overallocation of resources to the emitter industry.

In the liability case discussed here we assume that firms in industry 1 are compensated for damages and that potential entrants into the industry are aware that they too will be compensated. Firms in industry 2 including potential entrants are responsible for damages done to industry 1 and we assume, since the externality is diffuse and the firms are taken as identical, that each must bear the cost of compensation equally. With liability, profits for firms in industry 1 and 2 can then be written as:

$$(6) \quad \pi_1 = P_1 y_1 - C_1(y_1) - D_1(n_2 y_2) + [D_1(n_2 y_2)]$$

$$\pi_2 = P_2 y_2 - C_2(y_2) - [n_1 D_1(n_2 y_2)/n_2]$$

where the terms in brackets are compensation or liability payments by firms in each industry, respectively. The first-order conditions for maximum profits in each firm, assuming an interior solution, are

$$(7) \quad \partial\pi_1/\partial y_1 = P_1 - C_1' = 0$$

$$(8) \quad \partial\pi_2/\partial y_2 = P_2 - C_2' - n_1 D_1' = 0$$

which imply that the conditions for short-run optimality are satisfied.[9] However, if it is assumed that firms enter until profits are zero, compensation to industry 1 results in a long-run equilibrium (position L in Figure 1), the lowest point on the average cost curve without damages (AC_1^L) for firms in the receptor industry. Again the marginal cost function (C_1') does not shift since damages are separable and the resulting price (P_1^L) and output for each firm (y_1^L) under the liability case are less than optimum. Thus both *total* industry output and the number of firms will be too large for the receptor industry in

[9] A comparison of (7) and (8) with (2) and (3) will verify this statement when it is assumed that there is no entry or exit in the short run.

the liability case since demand for the industry's output will be greater at the lower price.

Turning to Figure 2, firms in industry 2 will reach an equilibrium point (L) in the long run under a liability rule which is the lowest point of the average total cost curve including each firm's share of damages to be paid (AC_2^L). Note that because total damages increase at an increasing rate, marginal damages (n_1D_1') are greater than average damages (n_1D_1/n_2y_2) so AC_2^L lies below AC_2^* and the intersection of the marginal social cost function ($C_2'+n_1D_1'$) and AC_2^L lies below and to the left of the optimum (*). This is the point of zero profits including liability for damages for the emitter industry. Both price (P_2^L) and output (y_2^L) are too low for each firm in industry 2. However, total industry output will be too high and there will be too many firms in the emitter industry under a liability rule since, given the lower price, aggregate industry demand will be too high. Thus, there results a long-run *overallocation* of resources to *both* industry 1 and 2. This sheds additional light on what Coase has called the Pigovian tradition, or in more modern form, the advocacy of taxation to correct external diseconomies by adjusting prices to equal marginal social costs. Taxation is necessary to prevent a misallocation of resources in long-run production even if there are well-defined property rights with respect to the externality such as in the liability case above. A liability solution *could* be adjusted to the optimum equilibrium point by taxing receptors an amount equal to the difference between marginal damages and average damages, a procedure sufficiently inefficient as regards information and enforcement costs to make a liability solution unattractive.

The nonliability case can best be explained in two steps. First, we will demonstrate that a negotiated solution under a nonliability rule cannot sustain the optimum points (*) assuming that the number of firms in each industry is constrained to be less than or equal to the optimum ($n_1 \leq n_1^*$, $n_2 \leq n_2^*$). As will be seen later, this assumption prevents a free rider problem from upsetting the potential equilibrium point (*) in Figures 1 and 2. Assume that firms in both industries are initially at *. We now observe that * for firm 1 in Figure 1 is a point of zero profits. However, in Figure 2 it is clear that firms in industry 2 could earn profits greater than those obtained under * by moving to point A. Thus, for * to be achieved by negotiation for firms in industry 2, firms in industry 1 must offer to pay a bribe at least equal to the difference between profits at * and profits at A to existing firms in industry 2. Clearly firms in industry 1 are making zero profits at * in Figure 1 and cannot pay the bribe. Thus, the optimum points * in Figures 1 and 2 are not feasible under a nonliability bargaining solution even ignoring the destabilizing effects of entry on coalitions since firms in industry 2 must be made at least as well off at market price P_2^* as they would be by not adjusting for the externality. Thus, firms in industry 2 would be unwilling to remain at *. It is conceivable, with the number of firms fixed by controlling entry through some licensing process, to achieve a short-run optimum with a solution somewhere along the marginal social cost function ($C_2'+n_1D_1'$) in Figure 2. However, the number of firms in industry 1 must be fixed ($n_1 \leq n_1^*$) such that profits sufficient to cover bribes to industry can be obtained. Clearly, without some taxation policy, even by controlling entry, the long-run optimum solution is not attainable under nonliability since receptors cannot initially afford to bribe emitters.

If free entry is allowed, potential entrants always have a valid threat of entry in the nonliability case if market prices are above P_2^E in industry 2 or above the lowest point of the current average cost plus average damages in industry 1. It is clearly impossible to bribe potential firms to stay out of an industry as long as they could earn earn positive profits by entering, because under perfect information regarding current profitability, an indefinitely large number of potential firms would eventually threaten to enter. In Figure 2, entry would result in an eventual price of P_2^E for firms in industry 2 with a short-run optimum position at point B, implying negative profits for emitter firms. However, this point cannot be stable be-

cause firms in industry 1 must earn sufficient profits to compensate firms in industry 2 for their losses. But free entry into industry 1 will tend to force profits in that industry to zero by a free rider process where receptors will enter (given a level of emissions reduced by negotiations between existing firms), join the coalition, but find profits eventually reduced to the point where firms in industry 2 can no longer be bribed to reduce output. Note that this process assumes that firms are price takers, i.e., perceived demand for each firm or potential entrant is infinitely elastic. Thus, potential entrants cannot realize that entry must lower price, destabilizing existing solutions. This sketch of events implies that under a nonliability rule with free entry negotiated solutions are unstable. One can again imagine a sufficiently complicated set of regulations and/or taxes to allow an optimal solution to be obtained under a nonliability rule, but these take on a Rube Goldberg character similar to the liability case.

Although we have excluded negotiation costs from this analysis, it is difficult to imagine any set of circumstances in which firms are competitive (free entry is of course a necessary condition for the existence of competitive firms), an externality exists between industries, and long-run optimality is achieved without taxation. The Coase proposition then reduces to an intuitive result. It may be preferable to allow courts to settle two-party disputes between private individuals over nuisances such as motorcycle noise by the allocation of rights. However, since it is a priori impossible to discover if a liability solution is Pareto superior to the unadjusted externality state in long-run production, policy makers would be well advised to rely on taxation or optimal standards where emitters bear the cost of meeting those standards.[10] Negotiations after taxation must involve a transfer of funds between industries and can be treated as collusive behavior.

II

So far, the analysis has implicitly assumed that the second-order conditions for a maximum of net benefits can be satisfied by a Pigovian tax and that the optimal long-run equilibrium point is dynamically stable. We will now demonstrate under quite general conditions that if the second-order conditions are satisfied, the optimal point is locally stable, and that under additional assumptions (essentially those underlying the previous comparative static analysis) local stability implies that the second-order conditions are satisfied. The latter is clearly the more significant in that it assures us a Pigovian tax is optimal if the dynamic equilibrium solution is an interior point. However, the optimal solution may be on the boundary, i.e., one of the industries might cease production. The possibility of boundary solutions should be considered because the presence of an externality will tend to make the long-run production frontier convex rather than concave. This implies that the second-order conditions are significant even where diminishing marginal productivity (increasing marginal cost) for each industry, taken separately, is assured. Further, it is also conceivable that only a local optimum will be achieved by corrective policies. (See Baumol and David Bradford.)

To assure at least the local optimality of an interior point, the first-order conditions (2)–(5) can be totally differentiated to give a matrix equation of the form shown as equation (9). The second-order conditions for a maximum of (1) require that the matrix A be negative definite or that $|A|>0$, $|A_{44}|<0$, $|A_{44.33}|>0$, and $|A_{44.33.22}|<0$.[11] Note that A (assuming the optimal tax is applied to correct for externalities that are present in production) will quite generally be a symmetric matrix ($A=A'$). In the specific case where damages are separable, the optimal tax on output of industry 2, (n_1D_1') can be calculated solely from the form of the damage function ($D_1(n_2y_2)$) and a knowledge

[10] An optimal standard where entry is considered must specify both allowable emissions for each firm and also regulate the number of emitters.

[11] The notation $|A_{hi.jk}|$ indicates the determinant of the matrix derived from A with rows h and j and columns i and k excluded.

$$(9)\qquad \begin{bmatrix} \\ A \\ \\ \end{bmatrix} \begin{bmatrix} dy_1 \\ dy_2 \\ dn_1 \\ dn_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where

$$A = \begin{bmatrix} n_1(n_1P_1' - C_1'') & 0 & n_1P_1'y_1 & 0 \\ 0 & n_2(n_2P_2' - C_2'' - n_1n_2D_1'') & -n_2D_1' & n_2y_2(P_2' - n_1D_1'') \\ n_1P_1'y_1 & -n_2D_1' & P_1'y_1^2 & -D_1'y_2 \\ 0 & n_2y_2(P_2' - n_1D_1'') & -D_1'y_2 & y_2^2(P_2' - n_1D_1'') \end{bmatrix}$$

---

of the number of firms in each industry. However, to check a priori if a true optimum is achieved by a solution of equations (2)–(5) we must be able to determine the elements of the A matrix. This clearly involves additional knowledge of the form of all cost functions and of the demand curves for each industry. Fortunately, an analysis of the dynamics of firm entry will allow us to dispense with this problem in the separable case and further demonstrate at least the theoretical validity of outcomes under Pigovian tax policies.

The dynamic model of entry applied here follows the work of Vernon Smith and others for the competitive case.[12] If we assume that output is adjusted instantaneously with the optimal tax applied, the conditions for a short-run profit maximum for n_1 firms in industry 1 and n_2 firms in industry 2 will be, respectively:

$$(10)\qquad n_1(P_1 - C_1') = 0$$

$$(11)\qquad n_2(P_2 - C_2' - n_1D_1') = 0$$

Profits for firms in each industry will then be

$$(12)\qquad \pi_1 = y_1P_1 - (C_1 + D_1)$$

$$(13)\qquad \pi_2 = y_2P_2 - (C_2 + n_1D_1'y_2)$$

This system of four equations can be solved to eliminate y_1 and y_2, so profits in each industry can be written as a function of the number of firms in both industries:

$$(14)\qquad \pi_1 = \pi_1(n_1, n_2) \qquad (\overset{*}{\pi}_1(\overset{*}{n}_1, \overset{*}{n}_2) = 0)$$

$$\pi_2 = \pi_2(n_1, n_2) \qquad (\overset{*}{\pi}_2(\overset{*}{n}_1, \overset{*}{n}_2) = 0)$$

Entry equations are assumed to take the form:

$$(15)\qquad \dot{n}_1 = H_1(\pi_1) \quad (H_1' > 0,\ H_1(0) = 0)$$

$$\dot{n}_2 = H_2(\pi_2) \quad (H_2' > 0,\ H_2(0) = 0)$$

where a dot above the variable indicates the time derivative. To determine if the system is stable about the equilibrium point (n_1^*, n_2^*), linearization proceeds as shown in equation (16).

$$(16)\qquad \begin{bmatrix} \dot{n}_1 \\ \dot{n}_2 \end{bmatrix} = \begin{bmatrix} H_1'(0)\partial\overset{*}{\pi}_1/\partial n_1 & H_1'(0)\partial\overset{*}{\pi}_1/\partial n_2 \\ H_2'(0)\partial\overset{*}{\pi}_2/\partial n_1 & H_2'(0)\partial\overset{*}{\pi}_2/\partial n_2 \end{bmatrix} \cdot \begin{bmatrix} n_1 - \overset{*}{n}_1 \\ n_2 - \overset{*}{n}_2 \end{bmatrix} + \begin{bmatrix} H_1(0) \\ H_2(0) \end{bmatrix}$$

Local stability is assured if the trace of the matrix of partial derivatives is negative,

$$(17)\qquad H_1'(0)\partial\overset{*}{\pi}_1/\partial n_1 + H_2'(0)\partial\overset{*}{\pi}_2/n_2 < 0$$

and if the determinant of that matrix is positive,

$$(18)\qquad H_1'(0)H_2'(0)(\partial\pi_1/\partial n_1\partial\pi_2/\partial n_2 - \partial\pi_1/\partial n_2\partial\pi_2/\partial n_1) > 0$$

Given the second-order conditions for a social optimum, the dynamic model of entry,

[12] For dynamic models of firm entry see Smith, E. Philip Howrey and Richard E. Quandt, and Myron G. Myers and E. Richard Weintraub.

and the condition for local stability as defined above we can state the following proposition:

PROPOSITION 1: *Given an objective criterion similar to that in equation (1) above (we can additionally allow interdependent demand functions, nonseparable externalities, and/or reciprocal externalities between the two industries) and assuming the first-order conditions (10) and (11) can be achieved by a Pigovian tax policy, the optimal equilibrium point will be locally stable if the second-order conditions are satisfied (i.e., the system is at a local optimum).*

Where output is adjusted instantaneously (9) can be rewritten as

$$\begin{bmatrix} & & & \\ & A & & \\ & & & \end{bmatrix}\begin{bmatrix} dy_1 \\ dy_2 \\ dn_1 \\ dn_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ d\pi_1 \\ d\pi_2 \end{bmatrix}$$

Solving this system for $\partial\pi_i/\partial n_j$ ($i, j=1, 2$) and applying Jacobi's Theorem on determinants[13] gives

(19)
$$\partial\pi_1/\partial n_1 = \frac{|A_{44}|}{|A_{44.33}|} \qquad \partial\pi_1/\partial n_2 = \frac{|A_{43}|}{|A_{44.33}|}$$
$$\partial\pi_2/\partial n_1 = \frac{|A_{34}|}{|A_{44.33}|} \qquad \partial\pi_2/\partial n_2 = \frac{|A_{33}|}{|A_{44.33}|}$$

Condition (17) can now be rewritten in terms of the A matrix by again applying Jacobi's Theorem to the second term in brackets as

$$(20) \quad H_1'(0)\left(\frac{|A_{44}|}{|A_{44.33}|}\right) + H_2'(0)\left(\frac{|A|}{|A_{44}|} + \frac{|A_{34}|\,|A_{43}|}{|A_{44.33}|\,|A_{44}|}\right)$$

By assumption $H_1'(0), H_2'(0)>0$ and since A is symmetric $|A_{34}|\ |A_{43}|>0$. Thus, if the second-order conditions are satisfied ($|A|>0$, $|A_{44}|<0$, $|A_{44.33}|>0$, $|A_{44.33.22}|<0$), expression (20) must be less than zero, satisfying the first of the stability conditions.

[13] Only one case of Jacobi's Theorem is needed throughout: $|A|\ |A_{44.33}|=|A_{33}|\ |A_{44}|-|A_{34}|\ |A_{43}|$.

Using (19) and applying Jacobi's Theorem once more, condition (18) can be rewritten in terms of the A matrix as:

$$(21) \qquad H_1'(0)H_2'(0)\frac{|A|}{|A_{44.33}|}$$

which by assumptions on $H_1'(0)$ and $H_2'(0)$ and fulfillment of the second-order conditions must be positive, satisfying the second condition for local stability.

This proposition implies that if the system is at or near a local optimum point it will tend to remain there, or more succinctly, that an optimum point is a point of stable equilibrium. Unfortunately, global optimality is more difficult to demonstrate.

PROPOSITION 2: *Given the objective criterion defined in (1) above and assuming the first-order conditions (10) and (11) can be achieved by a Pigovian tax if the following conditions are met: P_1', $P_2'<0$ and C_1'', C_2'', D_1', $D_1''>0$, then local stability implies local optimality.*

Note that Proposition 2 includes the assumption of separability on demand functions *and* on damages.

By inspection of the A matrix we note that all of the nonzero elements are negative by our assumptions on the signs of the relevant partial derivatives. This implies that $|A_{44.33.22}|<0$ and $|A_{44.33}|>0$. The signs of the higher order determinants can now be derived from the stability conditions.

From the second condition for local stability (21) we have

$$H_1'(0)H_2'(0)\frac{|A|}{|A_{44.33}|}>0$$

which implies that $|A|>0$ since $H_1(0)$, $H_2(0)$, $|A_{44.33}|>0$. The first condition for local stability (17) gives

$$\left(\frac{H_1'(0)}{|A_{44.33}|}\right)|A_{44}| + \left(H_2'(0)\left(|A| + \frac{|A_{34}|\,|A_{43}|}{|A_{44.33}|}\right)\right)\frac{1}{|A_{44}|}<0$$

which implies $|A_{44}|<0$ since the terms in brackets are positive by previous arguments.

This proposition assures us that, at least in separable cases, if a Pigovian tax policy yields a stable equilibrium point, then that point is a local optimum in the long run with free entry.

REFERENCES

W. J. Baumol, "On Taxation and Control of Externalities," *Amer. Econ. Rev.*, June 1972, *62*, 307–22.

——— **and D. F. Bradford**, "Detrimental Externalities and Non-Convexity of the Production Set," *Economica*, May 1972, *39*, 160–76.

D. F. Bramhall and E. S. Mills, "A Note on the Asymmetry Between Fees and Payments," *Water Resources Research*, Third Quarter 1966, *2*, 615–16.

G. Calabresi, "The Decision for Accidents: An Approach to Nonfault Allocation of Costs," *Harvard Law Rev.*, Feb. 1965, *78*, 713–30.

———, "Transactions Costs, Resource Allocation, and Liability Rules—A Comment," *J. Law Econ.*, Apr. 1968, *11*, 67–73.

R. H. Coase, "The Problem of Social Cost," *J. Law Econ.*, Oct. 1960, *3*, 1–44.

F. T. Dolbear, Jr., "On the Theory of Optimum Externality," *Amer. Econ. Rev.*, Mar. 1967, *57*, 90–103.

E. P. Howrey and R. E. Quandt, "The Dynamics of the Number of Firms in an Industry," *Rev. Econ. Stud.*, July 1968, *35*, 349–52.

A. V. Kneese and B. T. Bower, *Managing Water Quality: Economics, Technology, Institutions*, Baltimore 1968.

M. G. Myers and E. R. Weintraub, "A Dynamic Model of Firm Entry," *Rev. Econ. Stud.*, Jan. 1971, *38*, 127–30.

V. L. Smith, "Economics of Production from Natural Resources," *Amer. Econ. Rev.*, June 1968, *58*, 409–31.

R. Turvey, "On Divergences Between Social Cost and Private Cost," *Economica*, Aug. 1963, *30*, 309–13.

The Process Analysis Alternative to Statistical Cost Functions: Comment

By J. R. MARSDEN, D. E. PINGRY, AND A. WHINSTON*

Several recent articles have considered alternative techniques in characterizing production processes. Ken Smith, Marvin Miller, and Fred Golladay centered their attention on production of medical services while James Griffin applied his techniques to the petroleum industry. Each of these papers is concerned with short-run analysis characterizing currently existing and currently employed production processes. The purpose of this note is twofold: 1) to critically review Griffin's process analysis, and 2) suggest a generalized production theory alternative which overcomes several problems inherent in his process analysis. It is shown that Griffin's process analysis, while possibly having some computational advantages, suffers from many of the same difficulties which generate criticism for standard statistical fitting. Our generalized production theory alternative is based directly on engineering relationships and is designed to be capable of dealing with long-run planning problems. Further, as a special case, it is shown that properly placed restrictions on the optimization problem yields short-run results similar to Griffin's though not restricted by the assumptions of a linear programming framework. After a few short comments on data from existing plants and on standard statistical fitting techniques, Section I is devoted to a critical consideration of several of Griffin's claims for process analysis. Section II sketches a generalized production theory approach.

I

In order to compare the effectiveness of techniques for generating the production function, one must have an idea of the type of data which is available. Observations on existing plants are likely to have one or more of the following characteristics:

1) The observations will be closely grouped and only represent a single or few alternative technologies.

2) The firms will be using a nonproduction frontier technology because of additions or alterations to existing plants in response to relative price changes or because the firms are not profit maximizers.

3) The firms will be using a production-frontier technology but will be operating in the short run off the frontier.

A. *Statistical Technique*

Given the data restrictions, the best we can hope for this technique is good historical description. There are no short-run operational rules implied by a particular fitted functional form and there is no reason to expect that the fitted functional form will correspond to the production frontier. There is no solid basis for planning decisions nor can questions on such subjects as returns to scale or convexity of the production process be adequately handled. Even if the observed firms were operating on the frontier, the data would likely be very localized and extrapolation of any functional form would be extremely risky.

B. *Process Analysis*

Griffin's recent article proposed an alternative approach to statistical cost functions. It is our contention that Griffin's process analysis and the statistical cost function approach have basically the same disadvantages. Griffin makes the following claims for process analysis:

1) "This paper demonstrates the potential of the process analysis approach for *deriving* the short-run properties of cost curves. . . . In both cases [single and joint

* University of Kentucky, Virginia Polytechnic Institute and State University, and Purdue University, respectively. We wish to thank Kenneth Avio for discussions helpful in the development of this comment.

product] the results here *substantiate* the classical assumptions about short-run cost functions, i.e., that marginal costs slope upward and average costs are U-shaped" (p. 55, emphasis added).

2) " . . . changes in costs can be linked directly with the limited capacity of the process units and the substitution between the various processes" (p. 51).

3) " . . . the complete range of the cost curve can be investigated rather than a limited range of actual observations" (p. 51).

4) " . . . the effect of technological change as reflected in new processes and product mixes can be explicitly built into the cost function" (p. 51).

We take these claims in order:

1) It appears from these statements that Griffin views his petroleum industry example as a test of the classical cost function assumptions. The fact is that he obtains an increasing marginal cost curve as a direct implication of the convexity of the linear programming-activity analysis model (as he notes on p. 50). This is no different than assuming a Cobb-Douglas functional form for statistical estimation and then claiming an empirical result that substantiates the classical assumptions about the homogeneity of production functions. If one is to test the hypothesis that the cost curves are shaped in the classical manner, at the very least there should be a criteria to accept or reject the hypothesis. The question of convexity of the production process cannot be dealt with in Griffin's framework since he allows no possible alternatives. Griffin's upward sloping marginal cost curves do not imply convexity—they are an implication of his technique. The study of convexity requires the use of techniques such as those set out by S. N. Afriat and Giora Hanoch and Michael Rothschild for empirical production data. For the general production problem the technique suggested in our empirical paper is an appropriate one.

2) The marginal cost curve implied by the linear programming-activity analysis model is a step function (as noted by Griffin in fn. 10, p. 49) rather than a positively sloping continuous curve. The inverted L-shaped curve as discussed by Griffin is simply a special case where only one set of capacity constraints is binding over the range of the feasible output. The appropriate procedure for obtaining the marginal cost curve would be to use parametric programming and generate the step function over the feasible output range. Griffin, however, insists on connecting the points he generated with straight lines to further enhance his claim that his paper is a test of the classical cost function characteristics. His procedure eliminates the information which he claims the process analysis approach provides—the relation between the increasing marginal cost and the limited capacities. One could view Griffin's procedure as a first-order approximation of a convex non-linear technology. However, this does not eliminate the convexity assumption and its implication of an upward sloping marginal cost curve implicit in his approach.

3) Griffin's claim that the entire range of the cost function can be investigated is conditional on one's acceptance of the activity analysis formulation as a nonlocal representation. Activity analysis can be viewed as a first-order approximation of a set of non-linear production processes. The a_{ij} in this model do not have an engineering interpretation, but are simply the coefficients of the linear representation of the process. Since this is the case, then one would not expect the activity analysis formulation to be valid for levels of output some distance away from observed values.

4) The claim of the possibility of incorporating new technological processes into the cost function depends on being able to represent these new processes with activity analysis. Since the a_{ij} are not direct engineering coefficients but simply estimated coefficients obtained for a "representative plant," incorporating new technology is not possible unless a model of the system exists. If this is the case, why not use all of the information available instead of a first-order approximation?

The only apparent advantage of Griffin's

process analysis is a set of short-run operation rules for restricted output levels. The technique is not directly applicable to planning problems, nor can it be used to deal directly with questions relating to returns to scale or convexity of the production process. Further, it may even be true that fitting a non-linear statistical cost function will yield a more accurate estimation of cost than the linear activity analysis, although the former does not imply short-run operation rules.

II

As an alternative, consider some l-stage production process, K, acting on an mx1 input vector, I, yielding an nx1 output vector, O. Thus K is a mapping from m space to n space, the particular mapping being given by the exact specification of the process under consideration. If an exact specification of the process is achievable, then the mapping from I to O is fully determined.[1] It should be noted that the specification here is not an empirical or estimated relationship, but rather a hypothetical engineering specification. It is not limited to currently employed processes and is restricted only by engineering infeasibility. Complete specification of the process allows a selection of alternatives based on optimization criteria rather than a selection limited to some set of historically employed alternatives.

Given a complete specification of the process, a production frontier composed of maximal output levels obtainable from differing combinations of inputs is derivable. Each point, O_i^*, in this production frontier is associated with a particular technology which can be defined in terms of process sequencing or types of processes employed.

Mathematical approximation of the production frontier yields a production function for the process. No information is available from the production function that is not available directly from full specification of the process system. The production function is introduced for the purpose of compacting the system through the elimination of non-optimal alternatives and allowing a simplified tool for planning purposes. The production function is treated here as a convenient intermediate step in the determination of a general "least total cost" output relationship.

Now consider a division of the input variables into two groups—a division that may be accomplished based upon several criteria. For the purpose of this exposition, we simply base the division on "long-run" vs. "short-run" variables.[2] Once the division is completed, the optimization problems may be formulated as a two-stage non-linear programming problem. The non-linear problem can be solved using a non-linear algorithm such as the one described in Graves, Pingry, and Whinston.

More formally, using I_i to denote long-run inputs, a_i short-run inputs, we specify the problem as

$$\underset{I,a}{MAX}\, f(I_1, I_2, \ldots, I_p, a_1, a_2, \ldots, a_s) = \underset{I,a}{MAX}\, f(I, a)$$

$$\text{subject to} \qquad c \leq \bar{c}; \quad a \in A; \quad I \in \Omega;$$

where c represents input costs, $\bar{c}$ a possible cost constraint, and A and Ω possible constraints on the input sets a and I, respectively. The function $f(I, a)$ is based on engineering relationships, and implicit here are constraints such as mass balance, steady-state balance, and engineering restrictions.[3,4]

Viewing the problem in a two-stage framework yields the following form:

[1] A degree of randomness in a production process may preclude complete (exact) specification; i.e., certain random variables (such as growth rates of micro-organisms as discussed in our empirical paper) may be present and require adjustments in studying the process.

[2] This division of input variables is used here for purposes of comparison with Griffin's approach. Many alternatives are available. A detailed example of the application of generalized production theory is presented in the authors' paper where division of the input variables is based on variability of total cost with respect to each input's utilization.

[3] For a detailed example which explicitly considers such constraints, see the authors' paper.

[4] The approach outlined here is applicable to multiproduct cases using vector representations. This problem is largely ignored here because of the chosen direction of this note.

$$\underset{I}{MAX}\ \{\underset{a}{MAX}\, f(I, a)\}$$

subject to $\quad c \leq \bar{c}; \quad a \in A; \quad I \in \Omega$

The inner maximization problem will yield a function of the form $\bar{f}(I)$ enabling the problem to be rewritten as:

$$\underset{I}{MAX}\ \bar{f}(I)$$

subject to $\quad c \leq \bar{c}; \quad I \in \Omega$

Informally, a reciprocal of the two-stage problem is:

$$\underset{I}{MIN}\ \{\underset{a}{MIN}\ c(I, a)\}$$

subject to $\quad f(I, a) \geq \bar{0}; \quad a \in A; \quad I \in \Omega$

where $\bar{0}$ is some bound on output, and constraints such as those noted above (for example, material balance, steady-state balance, engineering restrictions) are again implicit in the formulation. The inner minimization problem in this formulation is a more general formulation of Griffin's process analysis but overcomes several of the disadvantages of that technique.

The nonconvex optimization technique has the following advantages:

1. It does not prejudge convexity and therefore does not necessarily imply the traditional rising marginal cost curve.
2. It allows changes in cost to be related to changes in technologies and capacities of various process elements.
3. It is based on engineering relationships and can be extended over the full range of cost curve.
4. It allows for the inclusion of new technologies using engineering relationships.
5. It provides an inner-optimization problem which is appropriate for deriving short-run operation rules which are independent of the validity of any convexity assumption.

The technique selected by an individual researcher depends on the problem faced. The statistical technique is most appropriate for a historical description. The process analysis approach is a short-run, first-order approximation which yields tentative short-run operation rules for convex technologies. This technique may prove a useful and inexpensive method to provide better short-run operation rules for specific plants. Application of the generalized production theory approach is more complex and hence more costly than the others. But it is the only technique broadly applicable to planning problems and to a detailed short-run analysis.

REFERENCES

S. N. Afriat, (1971a) "The Output Limit Function in General Convex Programming and Theory of Production," *Econometrica*, Mar. 1971, *39*, 309–39.

———, (1971b) "Efficiency Estimation of Production Functions," paper presented at the Aug. 1971 meetings of the Econometric Society.

G. W. Graves, D. E. Pingry, and A. Whinston, "Application of a Large Scale Nonlinear Programming Algorithm to Pollution Control," *Proc. Fall Joint Computer Conference*, Las Vegas 1971.

J. M. Griffin, "The Process Analysis Alternative to Statistical Cost Functions: An Application to Petroleum Refining," *Amer. Econ. Rev.*, Mar. 1972, *62*, 46–56.

G. Hanoch and M. Rothschild, "Testing the Assumptions of Production Theory: A Nonparametric Approach," *J. Polit. Econ.*, May/June 1972, Part II, *80*, 256–75.

J. R. Marsden, D. E. Pingry, and A. Whinston, "Production Function Theory and the Optimal Design of Waste Treatment Facilities," *Appl. Econ.*, Dec. 1972, *4*, 279–90.

K. Smith, M. Miller, and F. L. Golladay, "An Analysis of the Optimal Use of Inputs in the Production of Medical Services," *J. Hum. Resources*, Spring 1972, *7*, 208–25.

The Process Analysis Alternative to Statistical Cost Functions: Reply

By JAMES M. GRIFFIN*

Since my recent article, which cast considerable doubt on the statistical estimation approach to the derivation of cost functions, I have been expecting a comment from proponents of that approach. Instead, I am pleasantly surprised to be replying to disciples of the process analysis approach. J. R. Marsden, D. E. Pingry, and A. Whinston (MPW) feel that a linear programming application of process analysis to petroleum refining has basically the same disadvantages as the statistical cost function technique, but that these could be overcome with the adoption of their particular approach. They propose a more general formulation utilizing non-linear programming techniques and allowing for nonconvex production technologies. Let us begin by considering their four objections to my application of process analysis to petroleum refining.

First, MPW assert that the test of the classical cost function assumptions (i.e., marginal costs slope upward and average costs are U-shaped) was not really a test at all but proceeded directly from the convexity assumptions of the linear programming model of the refinery. Certainly, the fixed capital process constraints imply a finite output and a rising marginal cost curve, but the relevant question is over what output range do marginal costs rise. MPW apparently feel that because of the convexity assumption marginal costs must necessarily rise over a broad output range. To demonstrate the error in their assertion, one need only examine some output range from $b_j^{(n)}$ to $b_j^{(n+1)}$ over which the basis x^* does not change. Since the basis is unchanged, the dual solution vector y^* will similarly not change, thereby proving that short-run marginal costs (given by the jth element of y^*) are constant over the given output range. As an example, Figure 1 of my paper illustrates a case where a basis change did not occur over the output range 8.4 to 8.9 MMB/D and marginal costs are constant. Furthermore, the dots in Figure 1, indicating basis changes, suggest that even after basis changes, marginal costs need not necessarily increase as the basis changes. Therefore, under this standard linear programming problem where the production processes are convex, short-run marginal costs can either rise in a step-wise manner or remain horizontal over the output range up to the full utilization of the capital stock at which point marginal costs become vertical. Either rising short-run marginal costs or an inverted L-shaped short-run marginal cost can be obtained assuming a standard convex production technology. Since the same result may be found in statistical cost studies (i.e., constant short-run marginal costs over the observed output range), the results in both Figures 1 and 2 indicating a rising marginal cost function *over a broad output range* certainly do not follow from the convexity assumptions as MPW assert.

Secondly, MPW are apparently disturbed because the short-run marginal cost function as drawn in Figure 1 does not change in a step function manner. They argue that the use of parametric programming would have revealed these steps and other useful information regarding capacity limitations. Contrary to MPW's assertion, parametric programming with UNIVAC's *Omega* package was utilized which reports the activities entering and exiting the basis at each basis change. As indicated in footnote 10, page 49, the particular parametrics option chosen does not report the complete solution vector at each basis change within the δ increment to the output constraint b_j, but rather reports the solution values for the first basis change

* Department of economics, University of Pennsylvania and the University of Houston.

after output has been incremented *at least* by δ. Due to the fact that in some cases the basis would change ten times over a δ interval of .1 MMB/D (approximately 1 percent of output), this additional precision hardly seemed worth generating the additional printout. Therefore, because basis changes could occur within a δ output increment and thereby cause marginal costs to change, a simple linear interpolation was utilized between the full solution printouts as an approximation. This approximation in no way substantively affects the empirical results.

MPW's third concern is that in using a linear programming approximation to a nonlinear technology, the linear approximation may be appropriate for investigating the cost function only over a restricted output range. Although I do not feel it is a problem in the petroleum model, MPW have raised a very good point worth emphasizing. In addition to the issue of nonlinearities causing the technical coefficients (a_{ij}) to change, the researcher must avoid omitting technically feasible activities which are not used over existing output ranges and factor prices. For example, on the catalytic cracking process, the severity, temperature, and catalyst type can all be potentially varied, yet operation experience in particular refineries will be frequently limited to two or three activity vectors out of a possible fifteen process options. To omit these technically feasible but unused activity vectors can lead to substantial distortions when examining the cost function for extended ranges and alternative factor input prices. It was precisely for these reasons that the Bonner and Moore model was used, because this model included these alternative activity vectors in order to consider the deleading of gasoline. As for the distortion due to nonlinearities in the production process, the interaction between lead and octane is non-linear. As noted in footnote 5 and described in the Bonner and Moore model, this problem was treated as a separable programming problem (i.e., linearizing the non-linear segments), which enabled the problem to remain cast in a linear programming framework. Apparently, this treatment of nonlinearities is not adequate in MPW's view. It would certainly be interesting if MPW would provide a non-linear version of the Bonner and Moore model to give empirical support to their position.

Finally, MPW imply that the effects of technological change as reflected in new processes and product mixes cannot be built into the cost function except under very restrictive conditions. A refinery need not have installed a new process and observed its a_{ij} coefficients to introduce the effects of technological change; rather all that is required is that one of the large process developers furnish the a_{ij} coefficients. These coefficients are typically based on results from a pilot design, although they could presumably be based on the first-order Taylor series approximation to a set of equations which purport to describe the process. Of course, MPW feel that the omission of the higher order terms would cause much valuable information to be lost. Why then, if MPW's non-linear engineering equations describe new processes with such great accuracy, do process designers spend large sums in constructing and testing pilot versions of a new process? Could it be that despite their nonstochastic appearance, engineering equations have an implicit error term and are themselves only an approximation?

I agree with MPW that there are situations involving nonconvexities in which nonlinear optimization techniques, such as their algorithm, have a contribution to make. The reader should, however, be alerted to the fact that global maxima are not assured in such situations. The choice of techniques obviously depends on the problem at hand, and process analysis should not be limited to the application of only one technique. In the petroleum refining example, the technology is convex and even though there are nonlinearities these are amenable to separable programming techniques, which allow the problem to remain cast in a simple linear programming framework. In view of these considerations and the absence of any empirical evidence to support MPW's contentions that their extensions would alter my basic conclusions, there would appear to be no reason to adopt their approach.

REFERENCES

Bonner and Moore Associates, *U.S. Motor Gasoline Economics,* Vol. 1 and 2, Houston 1967, American Petroleum Institute, pub. 4002, 4003, Washington.

J. M. Griffin, "The Process Analysis Alternative to Statistical Cost Functions: An Application to Petroleum Refining," *Amer. Econ. Rev.,* Mar. 1972, *62,* 46–56.

Price Discrimination by Regulated Motor Carriers: Comment

By M. L. GREENHUT, M. J. HWANG, AND H. OHTA*

The purpose of this paper is to investigate theoretically the pricing interactions and similarities of related spatial markets. In particular, it examines alternative spatial demands and market conditions for a firm's product and determines the transport rates that would be charged by an independent carrier. The resulting delivered prices of the firm will then be viewed. Pursuant to the overall objective, Section I reviews the recent model presented by Josephine Olson in this *Review*. Section II generalizes that model in two respects: 1) the rigid assumption of constant demand elasticity is relaxed, and 2) monopolistic as well as competitive market conditions are examined. Section III derives and compares the spatial price schedules that would result under alternative conditions in the related commodity and shipper markets. Section IV concludes the paper.

I. A Transport Rate Model for (Motor) Carriers

Olson assumes the total demand for a shipper's product j at each mileage ring i from the shipper's plant is:

$$p_{ij} = (x_{ij}T_j)^{-1/\alpha_j} \tag{1}$$

where $\alpha_j > 1$ and price is inversely related to the number of shipments (x_{ij}) and the size of the shipment (T_j). She then equates $p(=MR)=MC$ under the extreme assumption *for a space economy* of perfect competition.[1] In particular, since MC is defined as the sum of the constant marginal cost of production c_j and the freight rate r_{ij}, the profit maximizing calculus leads to:

$$p_{ij} = c_j + r_{ij} \tag{2}$$

Though equation (1) actually provides a concave demand for the product, the implicit assumption of infinitesimally small shippers selling to each point in the market at a given price p enables the substitution of (2) into (1), and thus:

$$(x_{ij}T_j)^{-1/\alpha_j} = c_j + r_{ij} \tag{3}$$

$$x_{ij}T_j = (c_j + r_{ij})^{-\alpha_j} \tag{3'}$$

Equation (3′) may be considered as the demand for the motor carrier's service, so that the carrier's profit function may be defined as:

$$\Pi = \sum_i \sum_j (r_{ij}x_{ij}T_j - K_{ij}x_{ij}T_j) \tag{4}$$

where K_{ij} is the carrier's marginal cost of production (i.e., its cost in transporting the seller's good). This cost is assumed to increase at a constant rate with distance. Substituting (3′) into (4) *and* resorting again to the profit maximizing calculus establish:

$$\frac{r_{ij}}{K_{ij}} = \frac{\alpha_j}{\alpha_j - 1} + \frac{1}{\alpha_j - 1}\,\frac{c_j}{K_{ij}} \tag{5}$$

The r_{ij}/K_{ij} ratios, therefore, decline as distance increases. Olson (p. 398) accordingly claims that equation (5) implies the condition of spatial discrimination in price against *nearer buyers*.

It was demonstrated long ago in diverse (yet corresponding) ways in the literature on spatial price discrimination that under the demand conditions assumed in (1) above, the seller would discriminate in price against his distant (not nearer) buyers.[2] To see this,

* Professor of economics, Texas A&M University, and assistant professors, West Virginia University, and Aoyama Gakuin University, respectively. This paper is based on research funded by the National Science Foundation.

[1] See Greenhut (1974, ch. 3 ff) where it is shown that spatial oligopoly results when the cost of distance is significant, except in the extreme case where all buyers or sellers are located in the same place. Under other locational assumptions, the $p=MC$ possibility taken as given by Olson would require special assumptions.

[2] See Edgar Hoover, Arthur Smithies, Donald Dewey, and Greenhut (1956, pp. 50, 157–60).

substitute (5) into (2) to obtain:

$$p_{ij} = \frac{\alpha_j}{\alpha_j - 1}(c_j + K_{ij}) \tag{6}$$

It follows that $dp_{ij}/dK_{ij} = \alpha_j/(\alpha_j - 1) > 1$ since $\alpha_j > 1$ and thus discrimination takes place against *distant buyers*, contrary to Olson's claim. Indeed, direct reference to her profit maximizing freight rate equation, as set forth above in (5), or its variant (5′) below, reveals the misinterpretation:

$$r_{ij} = \frac{\alpha_j}{\alpha_j - 1} K_{ij} + \frac{1}{\alpha_j - 1} c_j \tag{5′}$$

Observe with respect to (5) and (5′) that even though the r_{ij}/K_{ij} ratio decreases monotonically with distance, the absolute difference between r_{ij} and K_{ij} is increasing, as should be clear in Figure 1.

Equation (5′), in other words, provides a straight line of slope $\alpha_j/(\alpha_j - 1)$ $[= \tan\theta' > 1]$. As distance increases, the difference between what the carrier charges and the actual freight cost rises. The discrimination is, therefore, actually directed against the distant buyers, and increases as distances increase. It is not the r_{ij}/K_{ij} ratio which explains the carrier's discrimination, but simply the slope of the r_{ij} curve resulting from changes in K_{ij}. It must be emphasized here that the subject discrimination is effected by the motor carrier and not by the seller of the product.

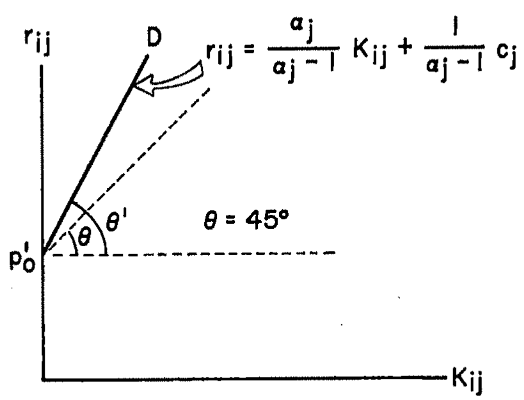


FIGURE 1

II. A Generalized Model and the Conditions Promoting Spatial Price Discrimination

The tendency for price discrimination to occur in economic space may be established more generally. Consider the following relationship where the assumption is continued that all buyers have identically the same gross demand curves (i.e., the same demand curves except for the cost of distance separating them from the seller), and where for notational simplicity we substitute c and D in place of c_j and r_{ij}, respectively.

From the condition $MR = MC$, obtain:

$$p\left(1 - \frac{1}{e}\right) = c + D \tag{7}$$

where
$$e = -\frac{dq}{dp}\frac{p}{q} \geq 1$$

By differentiating (7) with respect to D:

$$\frac{dp}{dD} = \frac{e}{\epsilon - (1 - e)} \tag{8}$$

where
$$\epsilon = \frac{de}{dp}\frac{p}{e}$$

This equation demonstrates sharply that:

a) $$\frac{dp}{dD} > 1 \Leftrightarrow 1 - e < \epsilon < 1$$

b) $$\frac{dp}{dD} < 0 \Leftrightarrow 1 - e > \epsilon$$

c) $$\frac{dp}{dD} = 1 \Leftrightarrow \epsilon = 1$$

d) $$0 < \frac{dp}{dD} < 1 \Leftrightarrow \epsilon > 1$$

It follows from a) that negative freight absorption implies and is implied by $1 - e < \epsilon < 1$, which includes the condition $\epsilon = 0$. Hence constant elasticity ($\epsilon = 0$) indicates negative freight absorption, which also means that mill prices rise as distance increases. Increasing elasticity of a specific type, $(1 - e < \epsilon < 0)$ as well as decreasing elasticity of a particular kind, $(0 < \epsilon < 1)$ will also lead to negative freight absorption. However, increasing elasticity does not al-

ways imply negative freight absorption. It could yield excessive freight absorption. In fact, result b) states that excessive freight absorption occurs if, and only if, increasing elasticity prevails of the type where $1-e>\epsilon$. In this case, however, the marginal revenue curve intersects the marginal cost from below and, accordingly, the intersection provides an unstable solution. Result c) indicates that no freight absorption occurs when elasticity decreases proportionately with price. In turn, observe from d) that positive freight absorption occurs if, and only if, elasticity decreases more rapidly than does price decrease, i.e., if, and only if, $\epsilon>1$. This last result relates, of course, to the linear and convex demand functions as well as to some concave demand functions in which elasticity also behaves as defined in d).

Recall now that in Olson's model the individual demand function $p_{ij}=(x_{ij}T_j)^{-1/\alpha_j}$ is assumed, and that α_j (the elasticity of demand) is given as a constant greater than unity. Constant elasticity in our model requires $\epsilon=0$, since $\epsilon=(de/dp)p/e$, and $de/dp=d\alpha/dp=0$. Contrary to Olson's claim, constant elasticity of demand promotes price discrimination against distant buyers in favor of proximate buyers.[3]

The sets of demand curves warranting discrimination against distant buyers are of rather special form. One might expect instead that typically $\epsilon>1$ in practice and, accordingly, that discrimination would proceed against nearer buyers. *Epsilons* of this order generate freight rate functions such as $p_o' D_o$ (see Figure 2).

To sum up: Olson's model involves $\epsilon=0$, hence discrimination would proceed against distant buyers. However, resale possibilities by nearer buyers militate against the use of this type of price discrimination. Moreover

[3] See supporting literature, fn. 2 above. And further recall that $q=x_{ij}T_j$. From the demand function $p=q^{-1/\alpha}$ and the definition of elasticity e as $-(dq/dp)p/q$, we see that $e=\alpha$. So, the elasticity of demand in Olson's demand function is constant at all points on the demand curve. If instead α was assumed to increase monotonically with increases in prices in conformance to the rule of *d*), discrimination against nearer buyers would have been deduced and statistical tests interpreted accordingly.

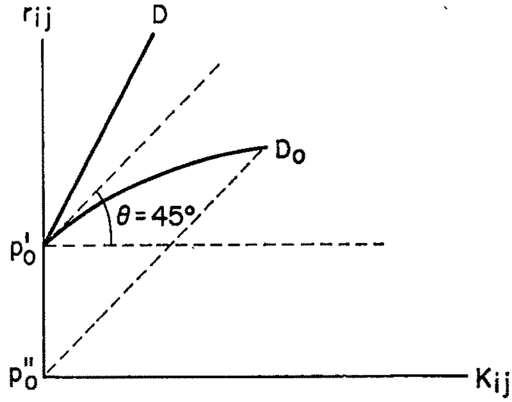


FIGURE 2

the assumption $\epsilon>1$ indicates that delivered prices increase *at a decreasing rate* as distance increases. Conceiving of $p_o'' D_o$ as the actual marginal cost in Figure 2, it follows that discrimination in the situation given by $\epsilon>1$ would be against proximate buyers in favor of distant buyers.

III. Related Markets, Derived Demand, and Alternative Spatial Price Schedules

It is, of course, an elementary proposition to observe that results obtained from a priori models depend upon the framework (assumptions) of that model. By utilizing less severe assumptions (i.e., by not assuming identical elasticities everywhere nor perfect competition), statistical results designed to test a given deduction can in turn be explained more precisely. At the same moment, the existence of price discrimination in the space economy can be better understood.

The model discussed previously in this paper is readily modifiable to conform to the relationships of Section II, and the framework proposed above. Assume, accordingly, the same gross individual demand curve for the shipping firm's product at each buying point (i.e., in each submarket) such that there exists:

$$p_{ij} = (x_{ij}T_j)^{-1/\alpha_j} \tag{9}$$

where, however, α_j is now a parameter the value of which increases with distance such that the condition $\epsilon>1$ applies. The monopo-

listic marginal revenue from the submarkets is readily derived as:

$$MR = p_{ij}\left(\frac{\alpha_j - 1}{\alpha_j}\right) \tag{10}$$

In order to maximize profits, MR must be equated with MC; this yields the profit maximizing p_{ij} and the output $x_{ij}T_j$ for each submarket, as given below:

$$p_{ij} = \frac{\alpha_j}{\alpha_j - 1}(c_j + r_{ij}) \tag{11}$$

$$x_{ij}T_j = \left(\frac{\alpha_j}{\alpha_j - 1}(c_j + r_{ij})\right)^{-\alpha_j} \tag{12}$$

Given $\alpha_j > 1$ and $\epsilon > 1$, the prices p_{ij} charged in the nearer markets are relatively higher than the prices in the distant markets when the marginal freight rates r_{ij} to each buying point are compared. As distance increases, the elasticity of demand α_j increases and $\alpha_j/(\alpha_j - 1)$ approaches unity. The price charged by the firm in turn reflects $MC = c_j + r_{ij}$ in the more distant markets, and at the limit market point the price p_{ij} becomes just equal to the marginal cost $c_j + r_{ij}$.

The amount of shipments of product j to each buying point i per unit of time, i.e., $x_{ij}T_j$, is smaller under conditions of imperfect competition than perfect competition since in the latter case $p = MC$. Under imperfect competition, $x_{ij}T_j$ is smaller at each submarket by the ratio $[(\alpha_j - 1)/\alpha_j]^{\alpha_j}$, where $(\alpha_j - 1)/\alpha_j < 1$ for $\alpha_j > 1$. Manifestly, the delivered price p_{ij} is higher and the output is lower under properly defined conditions of economic space than they are under the extreme assumption of perfect competition.[4] Thus, in Figure 3, $p_o''D_o$ would be the carrier's freight rate under conditions of perfect competition, and $p_o'D_o$ would be the freight rate of the monopolistic carrier which is discriminating in price. Assuming discriminatory freight rates, $\bar{p}_o\bar{D}_o$ would be the shipper's delivered price p_{ij} under conditions of perfect competition in the commodity market, and

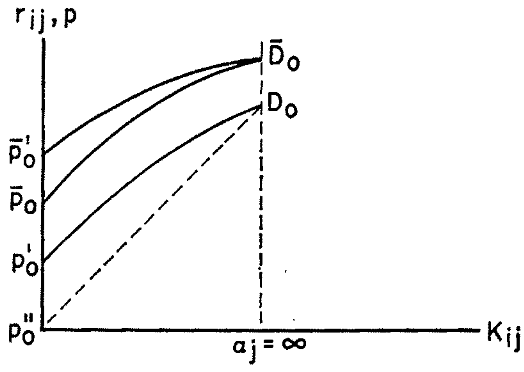


FIGURE 3

$\bar{p}_o'\bar{D}_o$ the delivered price under conditions of imperfect competition in the commodity market. Price discrimination in favor of distant buyers stems from the more elastic demand of these buyers, *ceteris paribus*.

It is further relevant to observe that r_{ij} in equation (12) is a component of the firm's marginal cost. But, it is also the delivered price from the standpoint of carriers. Therefore, equation (12) provides the demand function of the carrier, a different function as a result of the assumption of imperfect competition in the commodity market. (See (3′) which was based on perfect competition.) If the marginal revenue from the demand function is equated with marginal cost K_{ij}, the profit maximizing transport rate r_{ij} can be expressed as:

$$r_{ij} = \frac{\alpha_j}{\alpha_j - 1}K_{ij} + \frac{1}{\alpha_j - 1}c_j \tag{13}$$

in apparent conformity to equation (5′). Thus, given the constant marginal production cost, the profit maximizing freight rate of the motor carrier is a function of α_j (the elasticity of individual demand) and the actual marginal freight cost K_{ij}. *However, note that α_j is assumed here to be related to K_{ij}* and hence is no longer constant. It follows that because in the nearer market the elasticity of demand α_j is low but greater than 1, the factor $\alpha_j/(\alpha_j - 1)$ is relatively high; and the freight rate charged by the motor carrier is relatively high. As distance increases, the elasticity of demand α_j increases and

[4] It is neither realistic nor necessary to assume that a spatial commodity market is in perfect competition. In fact, it is almost impossible, if not theoretically inconceivable, to have perfect competition in a well-defined economic space, i.e., where freight cost is significant.

$\alpha_j/(\alpha_j-1)$ approaches unity. Moreover, the second term on the right-hand side of equation (13) declines and approaches zero.

Inserting the profit maximizing freight rate r_{ij} of the (motor) carrier, as given in equation (13), back into equation (11), yields:

$$(14) \quad p_{ij} = [\alpha_j/(\alpha_j - 1)]^2(c_j + K_{ij})$$

Thus, when both the seller of the product and his carrier possess monopolistic controls, the price discrimination practiced by shippers against nearer buyers is greater than when the discrimination is practiced on the level of the (motor) carrier industry alone (see Figure 3).

IV. Summary and Conclusions

Our arguments and findings may be completed as follows: Specifically, given the departure point of perfect competition in the (motor) carrier industry and in the commodity market, the firm's delivered price would be higher but parallel to $p_o''D_o$ in Figure 3, a straight line resulting in part from the assumption of a constant rising freight cost (K_{ij}) with respect to distance. When the unusual situation of perfect competition in the commodity market but monopoly in the motor carrier industry is imagined, the firm's delivered price is $\bar{p}_o\bar{D}_o$. When both markets are imperfect, its delivered price is $\bar{p}_o'\bar{D}_o$. Price discrimination is most pronounced in the last situation.

It should further be clear in light of reshipment (or resale) possibilities that spatial discrimination by a (motor) carrier (or shipper) tends not to be against the more distant buyers. Indeed, the (implicit) assumption by Olson of an ϵ that is not greater than unity rejects the demand functions one may expect to encounter in the real world. It is further proposed that the absolute difference between the carrier's profit maximizing freight rate r_{ij} and its actual freight cost explains price discrimination rather than the r_{ij}/K_{ij} ratio.

We also contend that if an underlying tendency exists to discriminate in price in favor of distant buyers in the commodity markets, carriers which fix their rates on the basis of value of service would lower their freight rates with greater distance. Motor carrier freight rate discrimination appears in this light as a derived demand result of commodity market price discrimination, and in effect involves monopoly pricing on the carrier level as a direct result of value of service pricing.

It is finally proposed that distant buyers may be expected to be favored by not only the concave downward rates of (motor) carriers, but also by the likely freight absorption practices of sellers in the commodity market. Indeed, since competition in economic space can be shown to be of imperfect order, our last proposition does not require simple monopoly on the shipper's level. This proposition and the basic market interdependencies described above indicate that the question of the overall impact on plant location of spatial pricing principles are additional related facets of a broad subject too long overlooked and unduly simplified in the past.[5]

[5] The simple theories of plant location which continue to play down or in fact to ignore the spatial interdependence and price practices of firms, in conformance with the Weberian cost approach, are surprisingly still prevalent today. In interesting contrast, see among other writings in recent years Nicos Devletoglou, David Smith, and Martin Beckmann.

REFERENCES

M. Beckmann, *Location Theory*, New York 1968.

N. Devletoglou, *Consumer Behavior*, New York 1971.

D. Dewey, "A Reappraisal of F.O.B. Pricing and Freight Absorption," *Southern Econ. J.*, July 1955, *22*, 48–54.

M. L. Greenhut, *A Theory of the Firm in Economic Space*, 2d ed., Austin 1974.

———, *Plant Location in Theory and in Practice*, Chapel Hill 1956.

E. M. Hoover, "Spatial Price Discrimination," *Rev. Econ. Stud.*, June 1937, *4*, 182–91.

J. Olson, "Price Discrimination by Regulated Motor Carriers," *Amer. Econ. Rev.*, June 1972, *62*, 395–402.

D. M. Smith, *Industrial Location*, New York 1971.

A. Smithies, "Monopolistic Price Policy in a Spatial Market," *Econometrica*, Jan. 1941, *9*, 63–73.

Price Discrimination by Regulated Motor Carriers: Reply

By JOSEPHINE E. OLSON*

M. L. Greenhut, M. J. Hwang, and H. Ohta have done an excellent job in pointing out the overly rigid assumptions of my model of motor carrier price discrimination (that is, constant elasticity of demand and perfect competition among shippers) and my incorrect interpretation of some of the results (that is, failing to recognize that the absolute difference between motor carrier rates and motor carriers costs increases with distance even though their ratio declines). Furthermore, they have developed a much more general model to explain why motor carrier discrimination varies with distance. If one assumes that elasticity of demand increases as the price of the good increases (a very logical assumption), then distant consumers will have more elastic demands and the difference between the profit maximizing motor carrier rate and motor carrier cost will decline.

There are only two points in their paper with which I wish to disagree. In Figure 1, they show that under my assumptions the absolute difference between motor carrier rates and motor carrier costs increases with distance, thus making the difference between price of good and cost of producing and shipping it greater in distant than in near markets. This fact leads them to state later that " . . . resale possibilities by nearer buyers militate against the use of this type of price discrimination." Resale of goods between two markets is profitable if the difference in price of the good between the two markets is greater than the additional shipping costs. Suppose there is a buyer at a short distance m from the original producer who is considering reselling to another buyer at a greater distance M, from the original producer. The distance between the two buyers is $M-m$. Assume that the motor carrier cost of shipping a good, K, is broken into its terminal and line-haul elements, K_o and m. Then the cost of hauling from the producer to the nearby buyer K_m, can be expressed as:

$$K_m = K_o + m \tag{1}$$

and for the long haul:

$$K_M = K_o + M \tag{2}$$

To rehaul the good from the nearby buyers to the distant one would cost:

$$K_{M-m} = K_o + M - m \tag{3}$$

Continuing the assumptions of my original paper, price in the distant market would be:

$$P_M = \frac{\alpha}{\alpha - 1}(c + K_o + M) \tag{4}$$

and price in the nearby market would be:

$$P_m = \frac{\alpha}{\alpha - 1}(c + K_o + m) \tag{5}$$

The difference in prices would be:

$$P_M - P_m = \frac{\alpha}{\alpha - 1}(M - m) \tag{6}$$

In order to resell the product, the nearby buyer must reship the good and thus the total transport rate will be higher than the rate for direct shipping to the distant buyer. The transport rates for the nearby buyer who wishes to resell will be:

$$r_m + r_{M-m} = \tag{7}$$

$$\frac{\alpha}{\alpha - 1}(K_o + m) + \frac{1}{\alpha - 1}c$$

$$+ \frac{\alpha}{\alpha - 1}(K_o + M - m) + \frac{1}{\alpha - 1}$$

* Assistant professor, business administration, Graduate School of Business, University of Pittsburgh.

The transport rate for a direct shipment to the distant buyer will be:

$$(8) \quad r_M = \frac{\alpha}{\alpha - 1}(K_o + M) + \frac{1}{\alpha - 1}c$$

The difference in transport rates is thus:

$$(9) \quad r_m + r_{M-m} - r_M = \frac{\alpha}{\alpha - 1}K_o + \frac{1}{\alpha - 1}c$$

Resale will be profitable only if:

$$(10) \quad \frac{\alpha}{\alpha - 1}(M - m) > \frac{\alpha}{\alpha - 1}K_o + \frac{1}{\alpha - 1}c$$

It can be seen that resale is not necessarily an imminent threat but tends to become a greater threat as the distance between the distant and nearby buyers increases. Because of the terminal element in motor carrier costs, greater discrimination against a distant buyer is still a possibility.

In an earlier version of my study (1969), I assumed that shippers as well as motor carriers were monopolistic. As Greenhut, Hwang, and Ohta show, this assumption makes no difference in terms of the rate structure of motor carriers but does mean that the price of the good in each market will be higher and the quantity sold less. The price in each market under the assumption of imperfect competition P_1, will be:

$$(11) \quad P_1 = \left(\frac{\alpha}{\alpha - 1}\right)^2 (c + K)$$

whereas the price under perfect competition P_2, is:

$$(12) \quad P_2 = \left(\frac{\alpha}{\alpha - 1}\right)(c + K)$$

Since α is greater than unity, P_1 is greater than P_2.

They then go on to conclude that "... when both the seller of the product and his carrier possess monopolistic controls, the price discrimination practiced by shippers against nearer buyers is greater than when the discrimination is practiced on the level of the (motor) carrier industry alone" (p. 784). This result follows directly from their assumption that price elasticity increases as the price of the good increases. If one retains my original assumption of constant elasticity, exactly the opposite conclusion is reached.

The increase in the imperfectly competitive price with distance is

$$(13) \quad \frac{dP_1}{dK} = \left(\frac{\alpha}{\alpha - 1}\right)^2$$

while the increase in the price under perfect competition is:

$$(14) \quad \frac{dP_2}{dK} = \frac{\alpha}{\alpha - 1}$$

Since α is greater than one, dP_1/dK is greater than dP_2/dK and there is relatively greater discrimination in the distant market under imperfectly competitive shippers than under perfectly competitive shippers.

REFERENCES

M. L. Greenhunt, M. J. Hwang, and H. Ohta, "Price Discrimination by Regulated Motor Carriers: Comment," *Amer. Econ. Rev.*, Sept. 1974, *64*, 780–84.

J. E. Olson, "Discrimination in Motor Carrier Class Rates," unpublished doctoral dissertation, Brown Univ. 1969.

———, "Price Discrimination by Regulated Motor Carriers," *Amer. Econ. Rev.*, June 1972, *62*, 395–402.

On the Durability of Capital Goods Under Imperfect Market Conditions

By WOLFHARD RAMM*

In a recent paper in this *Review*, Peter Swan showed that in the context of a simple model a monopolist would produce capital goods which are just as durable as those produced by a competitive firm having an identical cost structure.[1] This is a surprising result since most laymen and many economists have long suspected that monopoly power would lead a firm to produce goods which are less durable. In this paper it is argued that these suspicions are supported if some of Swan's assumptions are modified to reflect capital market conditions which appear to prevail in the markets for many capital goods. It is also found, however, that given plausible assumptions about purchaser expectations, a monopolist will find it profitable to produce a more durable good.[2]

I. Capital Market Imperfections

In this section I relax the Swan assumption that the rate at which users of the capital good discount future revenue and services equals the rate used by producers.[3] I retain the remaining assumptions of Swan: 1) The user of the capital good is concerned only with the flow of services it provides. 2) The flow of services of the capital good is proportional to its stock. 3) The good decays at a constant exponential rate which is a parameter chosen initially for all future production. 4) Per unit production costs are independent of the production rate and inversely related to the decay rate chosen by the producer.[4] 5) The rental rate on the services of the capital good is inversely related to the current supply of services. 6) The users and producers of the capital good have perfect foresight. 7) There is a perfect market for used units of the capital good.

The notation is summarized below. Subscripted variables denote partial differentiation with respect to the subscript variable.

r = the producer's discount rate; $r \geqq 0$
γ = the user's discount rate; $\gamma > r$
δ = the decay rate of the capital good
K = the stock of the capital good
p = the rental price of services provided by a unit of the capital good
P = the sale price of a unit of the capital good
C = the per unit cost of producing the capital good

The assumed demand relation for capital services may be written as

$$(1) \qquad p = p(K), \qquad p_K(K) < 0$$

* Assistant professor of economics, University of California, San Diego. I wish to thank Hiroshi Ono, Richard Schmalensee, and Peter Swan for their helpful comments on an earlier draft of this paper. Responsibility for errors and shortcomings is, of course, mine.

[1] Swan's paper was in response to earlier papers in which David Martin, E. Kleiman and T. Ophir, David Levhari and T. N. Srinivasan, and Richard Schmalensee (1970) show that a monopolist would produce a less durable good. Their work, however, contained an error which Swan corrected.

[2] Other attempts to relax the assumptions on which the Swan result is based include the introduction of capital market imperfections by Robert Barro, nonconstant costs by Morton Kamien and Nancy Schwartz, and maintenance possibilities by Schmalensee (forthcoming). Barro showed that if capital markets are imperfect, a monopolist may have an incentive to limit durability even though he could increase durability costlessly. We focus on explicit comparisons between monopolies and competitive firms in the more realistic case where unit costs rise as durability is increased.

[3] A divergence of discount rates appears to exist in the market for consumer durables, where one has in the past observed consumers borrowing at interest rates of 12 to 18 percent in order to finance durables, while the prime lending rate, presumably applicable to the producer, was 5 percent.

[4] This is another critical assumption in the analysis of durability. Kamien and Schwartz have shown that Swan's results are altered (the monopolist reduces durability) under the more realistic assumption of nonconstant short-run costs. Some of the results in this paper may also be altered if this assumption is relaxed.

since services are assumed to be proportional to the stock. The technological cost conditions are

$$(2) \qquad C = C(\delta), \qquad C_\delta(\delta) < 0$$

Under the assumptions of this section the optimal behavior of firms which rent the services of a capital good will differ from the behavior of firms which sell capital goods outright.[5] We consider first the case of firms which rent.

PROPOSITION 1: *If $\gamma > r$ and firms rent the services of capital goods, the durability of capital goods produced by competitive firms equals the durability of capital goods produced by monopolists.*

For a competitive firm, market equilibrium conditions require that the marginal costs of providing capital services equal the rental price of these services, which will equal marginal revenue under the assumptions of the competitive model. The marginal revenue to a firm from the sale of the services of an additional unit of capital maintained for a period of T years is

$$\int_0^T pe^{-rt}dt = p\frac{(1-e^{-rT})}{r}$$

The marginal cost of providing these services is

$$\int_0^T \delta C(\delta)e^{-rt}dt + [C(\delta) - C(\delta)e^{-rT}] = \frac{\delta C(\delta)(1-e^{-rT})}{r} + C(\delta)(1-e^{-rT})$$

where the first term represents stock replacement costs and the second term represents interest opportunity costs. Equating marginal revenue and marginal cost yields the condition

$$(3) \qquad p = (r+\delta)C(\delta)$$

Under the usual assumptions each firm will minimize the costs of providing capital services $[=(r+\delta)C(\delta)]$ with respect to δ. Differentiating these costs with respect to δ yields the following durability condition:

$$(4) \qquad \frac{1}{(r+\delta)} = -\frac{C_\delta}{C}$$

We consider next a monopolist who rents. He must select both the size of his stock, K, and the decay rate, δ. The present value of the monopolist's net revenues from the rental of his stock of the capital good are given by

$$(5) \qquad PV = \int_0^\infty [p(K)K - \delta C(\delta)K]e^{-rt}dt - C(\delta)K = \frac{p(K)K - \delta C(\delta)K}{r} - C(\delta)K$$

where the first term on the right-hand side of (5) represents the present value of net rental revenues and the second term represents initial stock production costs. First-order conditions for the maximization of PV require:

$$(6) \qquad PV_\delta = -\frac{K}{r}[C + (r+\delta)C_\delta] = 0$$

$$(7) \qquad PV_K = \frac{1}{r}[p_K K + p - (r+\delta)C] = \frac{1}{r}\left[\left(1-\frac{1}{E}\right)p - (r+\delta)C\right] = 0$$

where E is the absolute value of the price elasticity of demand for services of the capital good. Equation (6) can be rewritten as

$$(8) \qquad \frac{1}{(r+\delta)} = -\frac{C_\delta}{C}$$

[5] There are a number of reasons why a monopoly may prefer a sales policy, even if this results in the loss of the monopoly premium on the implicit loan services it provides when it rents. If the value of the capital good is low, the transactions costs involved in a rental arrangement may be prohibitive. If maintenance is necessary or the possibility of abuse exists, a moral hazard problem will exist. In fact, the characteristics which make a buyer a poor credit risk, and hence subject to a high borrowing rate, are likely to make a rental or lease arrangement risky for the seller of a capital good.

yielding a durability condition identical to that of a competitive firm.

A monopolist will, however, charge a higher rent than a competitive firm. The equilibrium rent p^{cr} of the competitive firm is given by (3). The rent a monopolist would charge p^{mr} is given, with some rearrangement, by (7):

$$p^{mr} = \frac{E}{(E-1)}(r+\delta)C \tag{9}$$

The monopolist's rent will be greater since monopoly profit maximization requires that $E>1$. Monopoly supply will be lower given the specified demand conditions.

Proposition 1 is not surprising since capital market imperfections do not affect firms that rent. If users do not purchase the capital goods, they do not have to make an intertemporal commitment. The rate at which users can borrow and lend should not, therefore, influence the conditions under which capital goods are produced.

One does, however, observe many capital goods being sold to users. We therefore consider next the case of a competitive firm which sells the capital good.

PROPOSITION 2: *If $\gamma>r$, a competitive firm which sells will make capital goods which are less durable than firms which rent the services of capital goods.*

In this case, the marginal revenue from the sale of a capital good will be equal to the sum of the future rents of a unit of the capital good discounted by the purchaser's discount rate:

$$P = \int_0^\infty pe^{-\delta t}e^{-\gamma t}dt = \frac{p}{\gamma+\delta} \tag{10}$$

where $e^{-\delta t}$ is the fraction of a unit of the capital good which remains after t years. Marginal costs will be equal to the constant per unit costs $C(\delta)$. Competitive equilibrium thus requires:

$$\frac{p}{(\gamma+\delta)} = C(\delta) \tag{11}$$

The competitive firm will attempt to maximize its per unit profits, $\pi^c = P - C$ with respect to δ. The first-order condition for the maximization of π^c with respect to δ yields, with some manipulation making use of (11), the durability condition:

$$\frac{1}{(\gamma+\delta)} = -\frac{C_\delta}{C} \tag{12}$$

Second-order conditions for profit maximization require:

$$2C_\delta + (\gamma+\delta)C_{\delta\delta} > 0 \tag{13}$$

Schmalensee (1970, p. 59) has shown that (13) implies that, in the case outlined above, the left-hand side of (12) will intersect the $-C_\delta/C$ curve from below.[6] The same line of reasoning also applies to the two rental cases considered in Proposition 1. Since $1/(\gamma+\delta) < 1/(r+\delta)$ by the hypothesis that $\gamma>r$, the left-hand side of (12) will intersect $-C_\delta/C$ at a lower level of durability than the left-hand side of (4) or (8), as is shown in Figure 1. We have thus shown that $\delta^{cr}=\delta^{mr}<\delta^{cs}$, where δ^{cr}, δ^{mr}, and δ^{cs} are the optimal durabilities of competitive firms that rent, monopoly firms that rent, and competitive firms which sell, respectively. The result of previous investigators that optimal durability is a decreasing function of the discount rate ($\delta_\gamma>0$) also applies in the three cases considered above.[7]

It is interesting to note that a competitive

[6] We consider explicitly only the monopoly sales case. We want to show:

$$\frac{d}{d\delta}\left[-\frac{C_\delta}{C}-\frac{1}{\gamma+\delta}\right] = -\frac{C_{\delta\delta}}{C}+\left[\frac{C_\delta}{C}\right]^2+\frac{1}{(\gamma+\delta)^2}$$
$$= -\frac{C_{\delta\delta}}{C}+\frac{2}{(\gamma+\delta)^2}<0$$

where use has been made of (11). The final inequality follows from the second-order condition (12). The same line of reasoning applies in the other two cases since the durability conditions and second-order conditions have the same form as those above.

[7] Differentiation of the durability condition (12) yields

$$\delta_\gamma = \frac{-C_\delta}{[2C_\delta + C_{\delta\delta}(\gamma+\delta)]} > 0$$

The numerator is positive since $C_\delta<0$ by assumption. Second-order conditions require the denominator to be positive. See Schmalensee (1970) or Swan.

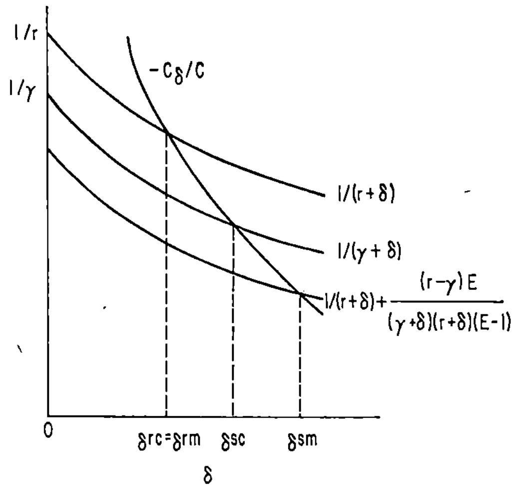


FIGURE 1

firm which rents the services of the capital good will charge a lower rent than a competitive firm which sells, and that a monopolist who rents may also provide services at a lower cost than the competitive seller. The competitive seller provides services for the implicit rental of

$$(14)\qquad p^{cs} = (\gamma + \delta)C(\delta)$$

which is an increasing function of the discount rate since

$$p_{\gamma}^{cs} = (1 + \delta_{\gamma})C + (\gamma + \delta)C_{\delta}\delta_{\gamma} = C > 0$$

where use has been made of (12). Thus p^{cs} is greater than p^{cr} since $r<\gamma$. We have already shown that a monopolist who rents will charge a higher rent than a competitive firm that rents. However, this difference is inversely related to the demand elasticity of the capital good. Taking the limit of (9) we find that

$$\lim_{E\to\infty} p^{mr} = (r + \delta)C = p^{cr}$$

Thus if E is sufficiently large, the monopoly rent will be close to that of the competitive renter and less than that of the competitive seller.

We consider next the monopolist who is constrained to sell the capital good.

PROPOSITION 3: *If $\gamma > r$, a monopolist who sells capital goods will make them less durable than a competitive firm which sells capital goods.*

The demand conditions faced by the selling monopolist are given by (10) since, as Swan has shown, the constant cost and perfect foresight assumptions of this paper imply that the optimal stock and implicit rental of the capital good are independent of time. Therefore the sale price is also independent of time since buyers will respond to attempts by the monopolist to vary the price of the capital good by imputing capital gains to ownership in the good, thereby holding the implicit rental constant. The present value of the monopolist's net revenues from sales of the capital good are thus given by

$$(15)\qquad \Pi = K\left[\frac{p(K)}{(\gamma+\delta)} - C(\delta)\right] + \int_0^{\infty}\delta K\left[\frac{p(K)}{(\gamma+\delta)} - C(\delta)\right]e^{-rt}dt$$
$$= K\left[\frac{p(K)}{(\gamma+\delta)} - C(\delta)\right] + \frac{\delta K}{r}\left[\frac{p(K)}{(\gamma+\delta)} - C(\delta)\right]$$

where the first term in (15) gives the net revenues from initial sales and the second term represents net revenues from replacement sales.

The first-order conditions for the maximization of present value are

$$(16)\qquad \Pi_K = \frac{1}{r}\left\{\frac{(r+\delta)}{(\gamma+\delta)}(p_K K + p) - (r+\delta)C\right\}$$
$$= \frac{1}{r}\left\{\frac{(r+\delta)p}{(\gamma+\delta)}\left[1 - \frac{1}{E}\right] - (r+\delta)C\right\}$$
$$= 0$$

$$(17)\qquad \Pi_{\delta} = \frac{K}{r}\left\{\frac{(\gamma-r)p}{(\gamma+\delta)^2} - C - (r+\delta)C_{\delta}\right\} = 0$$

Equation (16) can be solved for the implicit rental of the capital good,

$$p^{ms} = \frac{E}{(E-1)}(\gamma+\delta)C \tag{18}$$

The monopoly sales rental is thus greater than both the competitive sales rental, since $E>1$, and the rent charged by a monopolist who rents the good since $\gamma>r$. The implicit or explicit rental paid by the user of a capital good under alternative conditions can thus be ranked as follows:

$$p^{cr} < p^{cs} \gtrless p^{mr} < p^{ms}$$

Substitution of (18) into (17) and some algebra yields[8]

$$\frac{(r-\gamma)E}{(r+\delta)(\gamma+\delta)(E-1)} + \frac{1}{(r+\delta)} = -\frac{C_\delta}{C} \tag{19}$$

The left-hand side of (19) will always be smaller than the left-hand side of (12) since their difference equals

$$\frac{(\gamma-r)}{(r+\delta)(\gamma+\delta)(E-1)} > 0$$

where the inequality follows from the condition that $E>1$. Thus, as is shown in Figure 1, the left-hand side of (19) will always intersect $-C_\delta/C$ at a lower level of durability than was the case for the competitive firm which sells.

It is clear from (19) that the optimal level of durability is a function of the difference, $e=\gamma-r$, between producer and user discount rates. We investigate next the effect of changes in the configuration of these rates on monopoly output and durability.

PROPOSITION 4: *An increase in r, holding constant γ, will result in an increase in monopoly output and an increase in the durability of the output.*

Differentiating (16) and (17) totally yields the following system:

$$\begin{bmatrix} \Pi_{\delta\delta} & \Pi_{\delta K} \\ \Pi_{K\delta} & \Pi_{KK} \end{bmatrix} \begin{bmatrix} d\delta \\ dK \end{bmatrix} = \begin{bmatrix} \Pi_{\delta r} & \Pi_{\delta\gamma} \\ \Pi_{Kr} & \Pi_{K\gamma} \end{bmatrix} \begin{bmatrix} -dr \\ -d\gamma \end{bmatrix} \tag{20}$$

The desired partial derivatives can now be obtained by dividing (20) by dr, setting $d\gamma=0$, and solving the system for δ_r and K_r using Cramer's rule:

$$\delta_r = \frac{-\Pi_{\delta r}\Pi_{KK} + \Pi_{Kr}\Pi_{\delta K}}{H} < 0 \tag{21}$$

$$K_r = \frac{-\Pi_{Kr}\Pi_{\delta\delta} + \Pi_{\delta r}\Pi_{K\delta}}{H} > 0 \tag{22}$$

where H is the determinant of the matrix of partial derivatives on the left side of (20). The inequalities in (21) and (22) follow from the second-order conditions for the maximization of (15) which require $\Pi_{\delta\delta}<0$, $\Pi_{KK}<0$, $H>0$, and the signs of the cross-partial derivatives which are established below:

$$\Pi_{K\delta} = \Pi_{\delta K} = \frac{1}{r}\left[\frac{(\gamma-r)p(1-1/E)}{(\gamma+\delta)^2} - C - C_\delta(r+\delta)\right] < 0$$

(since $\Pi_\delta=0$, and it is easy to show that $\Pi_{K\delta}<\Pi_\delta$),

$$\Pi_{\delta r} = \frac{K}{r^2}\left[\frac{-\gamma p}{(\gamma+\delta)^2} + C + \delta C_\delta\right] = \frac{-pK}{r(\gamma+\delta)(r+\delta)E} < 0$$

(where use has been made of (17) and (18)),

$$\Pi_{Kr} = \frac{-\delta}{r^2}\left[\frac{p(1-1/E)}{(\gamma+\delta)} - C\right] = 0$$

(since $C=[p(1-1/E)]/(\gamma+\delta)$ by condition (16)).

The result obtained in (21) is surprising. It indicates that an increase in the monopo-

[8] With some algebraic manipulation one can show this relation to be equivalent to one displayed by Barro (fn. 7, p. 601).

list's discount rate will result in an increase in the optimal durability of capital goods produced by the monopolist if users' discount rates remain constant. In the perfect capital markets case, an increase in discount rates results in a decrease in durability. Equation (22) indicates that the monopolist will also increase the size of the stock he produces.

Some insight into these results can be obtained by substituting $\gamma - e$ for r in the present value function (15). The discount rate gap, e, then appears only in the term $\delta K/(\gamma - e)$, which represents replacement sales. A decrease in e, resulting from the increase in r, thus decreases the present value of replacement sales relative to the value of initial sales. The monopolist adjusts by making the good more durable, thereby placing a greater emphasis on initial sales. Since this also tends to reduce the cost of capital to the buyer, an increase in the optimal stock of the capital good results.

We next establish:

PROPOSITION 5: *An increase in γ, holding constant r, will result in a decrease in monopoly output and a decrease in the durability of the output, provided $(\gamma - r) < (r + \delta)$.*[9]

Our line of reasoning is the same as that under Proposition 4. We divide (20) by $d\gamma$, set $dr = 0$ and solve the system obtaining:

$$(23) \quad \delta_\gamma = \frac{-\Pi_{\delta\gamma}\Pi_{KK} + \Pi_{K\gamma}\Pi_{\delta K}}{H} > 0 \quad \text{if } (\gamma - r) < (r + \delta)$$

$$(24) \quad K_\gamma = \frac{-\Pi_{K\gamma}\Pi_{\delta\delta} + \Pi_{\delta\gamma}\Pi_{K\delta}}{H} < 0 \quad \text{if } (\gamma - r) < (r + \delta)$$

The inequalities follow since

$$\Pi_{\delta\gamma} = \frac{K(\delta + 2r - \gamma)p}{r(\gamma + \delta)^3} > 0 \quad \text{if } (\gamma - r) < (r + \delta)$$

$$\Pi_{K\gamma} = \frac{-(r + \delta)p(1 - 1/E)}{r(\gamma + \delta)^2} < 0$$

[9] The size of the gap between discount rates, $\gamma - r$, is also restricted in Proposition 4 since we assume that $r \geqq 0$.

The signs of the remaining partial derivatives and H were established following Proposition 4.

The interpretation of Proposition 5 is similar to that of Proposition 4. In this case the value of initial sales declines relative to replacement sales. With increased discount rates buyers also discount future services more, giving the monopolist an added incentive to reduce durability.

Finally, we examine the response of the monopolist to a general rise in the level of discount rates, defined as an increase in r holding e constant.[10]

PROPOSITION 6: *A general increase in discount rates results in a decrease in monopoly output and the durability of the output if e is sufficiently small.*

The monopolist's present value function can be expressed in terms of e and r as follows:

$$(25) \quad \bar{\Pi} = K\left[\frac{p(K)}{(r + \delta + e)} - C(\delta)\right] + \frac{\delta K}{r}\left[\frac{p(K)}{(r + \delta + e)} - C(\delta)\right]$$

The desired partial derivatives are computed in the same manner as under Proposition 4, yielding:

$$(26) \quad \delta_r = \frac{-\bar{\Pi}_{\delta r}\bar{\Pi}_{KK} + \bar{\Pi}_{Kr}\bar{\Pi}_{\delta K}}{\bar{H}} > 0$$

$$(27) \quad K_r = \frac{-\bar{\Pi}_{Kr}\bar{\Pi}_{\delta\delta} + \bar{\Pi}_{\delta r}\bar{\Pi}_{K\delta}}{\bar{H}} < 0$$

The inequalities follow, if the restrictions on e are satisfied, since

$$\frac{r^2}{K}\bar{\Pi}_{\delta r} = \frac{-2rep}{(r + \delta + e)^3} - rC_\delta > 0 \quad \text{if } e < \frac{-rC_\delta(r + \delta + e)^3}{2rp}$$

[10] This is not, of course, the only manner in which a general increase in the discount rates can be defined. One could require, for example, that percentage changes in both be the same.

(where use has been made of $\Pi_\delta = 0$)

$$r^2 \overline{\Pi}_{Kr} = \frac{-p(1 - 1/E)[(r+\delta)^2 + e\delta]}{(r+\delta+e)^2} + \delta C < 0$$

$$\text{if } e < \frac{r(r+\delta)}{\delta}$$

The remaining signs; $\overline{\Pi}_{\delta K} = \overline{\Pi}_{K\delta} < 0$, $\overline{\Pi}_{\delta\delta} < 0$, $\Pi_{KK} < 0$, and $\overline{H} > 0$ are readily established using the line of reasoning developed in Proposition 4.

The result in (26) generalizes the perfect capital market (i.e., $e = 0$) result that an increase in discount rates lowers durability. The result does not appear to hold, however, if e is large relative to r. When this is the case an identical increase in both r and γ will decrease the monopolist's revenues from replacements much more than it reduces purchasers' valuations of future services. In this case the monopolist may find it profitable to increase durability, thereby shifting some of his replacement sales to initial sales.

II. Purchasers with Stationary Expectations

Since businessmen and consumers clearly do not have perfect information about future price changes when they purchase capital goods, the behavior of a monopolist facing buyers with imperfect information is analyzed in this section. The assumption of perfect foresight is replaced with the assumption that buyers of the capital good have stationary expectations about the level of prices.[11] It is also assumed that capital markets are perfect (i.e., $\gamma = r$) and, with little loss of generality, that $p(K)$ has an upper bound. The remaining assumptions and notation of Section I are retained.

The effect of alternative expectations assumptions on the sales policy of a monopolist has been analyzed by A. J. Douglas and S. M. Goldman, who found that the monopolist would produce a larger stock of the durable good and charge a lower steady-state implicit rental under a broad range of expectations assumptions. Ronald Coase, in considering strategic aspects of selling a durable good with a zero decay rate (i.e., land), argues that a monopolist will find it profitable to expand the supply of land on the market to the competitive level if the marginal buyer has stationary expectations. K. G. Löfgren obtained a similar result.

The Douglas-Goldman and Coase results can be explained by dynamic price discrimination.[12] A monopolist will always find it profitable to extract as much consumers' surplus from a market as his ability to separate markets permits. Stationary expectations allow him to separate his markets as finely as he likes. When a new product is introduced to a market he can set its price so that just one unit is purchased. He can then lower the price just enough to sell the second unit. By continuing this process, the monopolist could capture all of the consumers' surplus available on initial sales. In the case of goods which decay, however, he will stop short of reducing his price to the competitive level since he has to balance his gains from discrimination on initial sales against losses of his monopoly premium on replacement sales in deciding on a steady-state stock.

Dynamic price discrimination can be modeled as follows. We assume that the monopolist builds up his stock so that $dK/dt = \alpha$, a constant which he can choose as large as he likes, always charging the maximum amount that buyers are willing to pay for an incremental unit, until he reaches his optimal steady-state stock. He then sells replacements to maintain that stock forever. The present discounted value of his sales are thus given by equation (28). The first term represents net revenues from the initial stock buildup, the second term represents replacement sales during the stock buildup, and the last term represents steady-state replace-

[11] Both the assumption of perfect foresight and stationary expectations are extreme and probably untenable in the case of most buyers. However, the line of reasoning developed in this section should also apply to other expectations mechanisms which underpredict actual price changes. We therefore consider only the analytically manageable stationary expectations case.

[12] See Löfgren for an earlier discussion of dynamic price discrimination.

ment sales.

$$(28)\quad R^* = \int_0^{K/\alpha} \alpha\left[\frac{p(\alpha t)}{(r+\delta)} - C(\delta)\right] e^{-rt}dt + \int_0^{K/\alpha} \alpha\delta t\left[\frac{p(\alpha t)}{(r+\delta)} - C(\delta)\right] e^{-rt}dt + \int_{K/\alpha}^{\infty} \delta K\left[\frac{p(K)}{(r+\delta)} - C(\delta)\right] e^{-rt}dt$$

Given the constant cost assumption of this paper, per unit costs will be independent of the monopolist's choice of α. It is therefore clearly optimal for the monopolist to obtain his revenues from initial sales as quickly as possible by making α arbitrarily large.[13] If a monopolist does follow such a policy, the function (28) can be simplified considerably. Taking the limit of (28) as α becomes arbitrarily large we obtain:[14]

$$(29)\quad \lim_{\alpha\to\infty} R^* = \lim_{\alpha\to\infty} \int_0^{K/\alpha} \frac{p(\alpha t)\alpha}{(r+\delta)} e^{-rt}dt - C(\delta)K + \frac{\delta K}{r}\left[\frac{p(K)}{(r+\delta)} - C(\delta)\right]$$

We can therefore approximate (28) by

$$(30)\quad R = \int_0^{K/\alpha} \frac{p(\alpha t)\alpha}{(r+\delta)} e^{-rt}dt - C(\delta)K + \frac{\delta K}{r}\left[\frac{p(K)}{(r+\delta)} - C(\delta)\right]$$

[13] The assumption that α can be made arbitrarily large is of course unrealistic. It is however an implication of the Swan assumptions.

[14] Equation (28) can be rewritten as

$$R^* = \int_0^{K/\alpha} \frac{p(\alpha t)\alpha}{(r+\delta)} e^{-rt}dt - \int_0^{K/\alpha} C\alpha e^{-rt}dt + \int_0^{\infty} \delta K\left[\frac{p}{(r+\delta)} - C\right] e^{-rt}dt + \int_0^{K/\alpha} \alpha\delta t\left[\frac{p(\alpha t)\alpha}{(r+\delta)} - C\right] e^{-rt}dt - \int_0^{K/\alpha} \delta K\left[\frac{p}{(r+\delta)} - C\right] e^{-rt}dt$$

To obtain (29) retain the first term. Replace p in the last two terms by its least upper bound. Then integrate the last four terms and take the limit $\alpha\to\infty$ of the terms involving α, applying l'Hôpital's rule as required. The final two terms vanish leaving (29).

in the analysis which follows with little loss of generality since

$$\lim_{\alpha\to\infty} R^* = \lim_{\alpha\to\infty} R$$

First-order conditions for the maximization of (30) require:

$$(31)\quad R_K = \frac{p(K)e^{-rK/\alpha}}{(r+\delta)} + \frac{\delta p(K)}{(r+\delta)r} + \frac{\delta K p_K}{(r+\delta)r} - \frac{(r+\delta)C}{r} = 0$$

$$(32)\quad R_\delta = \int_0^{K/\alpha} \frac{-p(\alpha t)\alpha}{(r+\delta)^2} e^{-rt}dt + \frac{K}{r}\left[\frac{rp}{(r+\delta)^2} - C - (r+\delta)C_\delta\right] = 0$$

If we let $\alpha\to\infty$ in (31) we obtain with some algebra:

$$(33)\quad C = \frac{p + \delta K p_K/(r+\delta)}{(r+\delta)}$$

which is equivalent to the steady-state condition Douglas and Goldman (equation (17)) obtained for a monopolist selling to consumers with stationary expectations. If we assume $\delta = 0$ in (33), we obtain the Coase and Löfgren result: $P = p/r = C$ which is the usual equilibrium condition for a perfectly discriminating monopolist. This result is not surprising since if the capital good does not decay, the monopolist captures the consumers' surplus on all of the present and future discounted services of the good.

With respect to the choice of durability we show:

PROPOSITION 7: *If purchasers of the capital good have stationary expectations, a monopolist will produce a more durable good than a competitive firm.*

Equation (32) can be written in the form of previous durability conditions as follows:

$$(34)\quad \frac{r}{CK(r+\delta)^3}\left[\int_0^{K/\alpha} p(\alpha t)\alpha e^{-rt}dt - pK\right] + \frac{1}{(r+\delta)} = -\frac{C_\delta}{C}$$

Using the same line of reasoning as was used in Propositions 1–3, it is clear that (34) implies that the monopolist will produce a more durable good only if the first term on the left side of (34) is positive. In this case the monopoly curve would be everywhere above the $1/(r+\delta)$ curve in Figure 1.[15]

The first term on the left side of (34) will be positive if the term inside brackets is positive since C, K, r, and δ are all positive by definition. To see that the term inside the brackets will always be positive consider first the case when demand is perfectly elastic, i.e., $p(K)=p^0$ for all K. Substituting, integrating, and taking the indicated limit we obtain:

$$\lim_{\alpha\to\infty}\int_0^{K/\alpha} \alpha p^0 e^{-rt}dt - p^0K = p^0K - p^0K = 0$$

Thus, if demand is perfectly elastic, condition (34) reduces to the competitive durability condition. However, if demand is less than perfectly elastic and monotonically decreasing in αt, we have

$$(35)\quad \lim_{\alpha\to\infty}\int_0^{K/\alpha} p(\alpha t)\alpha e^{-rt}dt > p\left(\alpha\frac{K}{\alpha}\right)K = p(K)K$$

since the left side of (35) includes all the consumers' surplus from the initial sale of K units while the right side is equal to the revenue which could be obtained from the sale of K units without price discrimination. This establishes the proposition since this consumers' surplus will always be positive under assumed demand conditions.

Proposition 7 can be explained as follows. A normal monopolist, one who faces consumers with perfect foresight, has to be content with normal monopoly profits on both initial sales and replacement sales. A dynamically discriminating monopolist, however, earns an additional premium from price discrimination on his initial sales. He therefore has an incentive to make his goods more durable in order to shift the sale of services from replacement sales to initial sales, thereby increasing the quantity of services which are sold subject to price discrimination.

III. Conclusion

The results of this paper indicate that the relationship between market structure and a capital good producer's choice of durability will depend on the conditions under which the good is produced. In the Swan model demand conditions, and hence market structure, do not affect the durability decision of a monopolist since the marginal gross revenue gain on initial sales due to an increase in durability will just equal the marginal gross revenue loss on replacement sales. This symmetry between initial sales and replacement sales breaks down, however, when capital market imperfections exist since increments to initial sales become less valuable. Stationary expectations, on the other hand, make increments to initial sales more valuable. In both cases the monopolist adjusts to the asymmetry by altering durability, thereby shifting the sale of services between initial and replacement sales.[16] Whether a monopolist will produce more, or less, durable goods than a competitive firm thus appears to depend on general market conditions.

[15] It should be noted that the durability chosen by a firm in a competitive industry will not depend on buyer expectations. Since a competitive firm can not influence its current price or future sales by assumption, it will maximize its profits by minimizing per unit costs at each point in time. The durability condition (4) is thus also appropriate for a competitive firm facing buyers with stationary expectations.

[16] A similar interpretation can also be given the results of Kamien and Schwartz. In their model attempts by the monopolist to build up the stock rapidly increase unit costs, since the scale of the plant is fixed in the short run, thereby reducing profits on initial sales. The monopolist therefore makes a less durable good, thereby deemphasizing initial sales and spreading the sale of services more evenly over time.

Our results are also subject to some qualification. The models which we analyzed contain a number of simplifications which are not particularly defensible. In particular, the assumption that decay rates must be chosen initially for all future time is clearly too strong. In the case of a dynamically discriminating monopolist, for example, it may be optimal for the monopolist to reduce the durability of his output once he has maximized his revenues from initial sales. Similar considerations may also apply to the monopolist facing capital market imperfections. As we have already noted, the constant, short-run cost assumption is also clearly unrealistic.

REFERENCES

R. J. Barro, "Monopoly and Contrived Depreciation," *J. Polit. Econ.*, May 1972, *80*, 598–602.

R. H. Coase, "Durability and Monopoly," *J. Law Econ.*, Apr. 1972, *15*, 143–50.

A. J. Douglas and S. M. Goldman, "Monopolistic Behavior in a Market for Durable Goods," *J. Polit. Econ.*, Jan./Feb. 1969, *77*, 49–59.

M. I. Kamien and N. L. Schwartz, "Product Durability Under Monopoly and Competition," *Econometrica*, forthcoming.

E. Kleiman and T. Ophir, "The Durability of Durable Goods," *Rev. Econ. Stud.*, Apr. 1966, *33*, 165–78.

D. Levhari and T. N. Srinivasan, "Durability of Consumption Goods: Competition versus Monopoly," *Amer. Econ. Rev.*, Mar. 1969, *59*, 102–09.

K. G. Löfgren, "The Theory of Intertemporal Price Discrimination. An Outline," *Swedish J. Econ.*, Sept. 1971, *73*, 333–43.

D. D. Martin, "Monopoly Power and the Durability of Durable Goods," *Southern Econ. J.*, Jan. 1962, *28*, 271–77.

R. Schmalensee, "Market Structure, Durability, and Maintenance Effort," *Rev. Econ. Stud.*, forthcoming.

———, "Regulation and the Durability of Goods," *Bell J. Econ.*, Apr. 1970, *1*, 54–64.

P. L. Swan, "Durability of Consumption Goods," *Amer. Econ. Rev.*, Dec. 1970, *60*, 884–94.

Commodity Trade and Factor Mobility

By MELVYN B. KRAUSS*

Since Paul Samuelson's factor-price equalization theorem that commodity trade could serve as a perfect substitute for factor mobility, the "mirror image" problem of the conditions under which factor mobility could perfectly substitute for commodity trade has attracted the attention of such trade theorists as Robert Mundell, Ernest Nadel, and most recently Frank Flatters. The latter, in analyzing the case where factor-owners move with their exported productive factors, does not consider the question of the uniqueness of the equilibrium obtained. This is the subject of the present paper. It will be demonstrated that in the case where factor prices differ between countries in closed economy equilibrium, and where such difference is due to different factor endowment ratios in the two countries, factor-price equalization is consistent with an infinite number of combinations of factor flows, the determination among which on a priori grounds is impossible. There will, however, be a single determinate "mirror image" solution in the case where trade is due to differences in tastes between the two countries.

I

The framework of the analysis is the "two by two by two" model of international trade (two factors, two commodities, and two countries) in which good X is unambiguously defined as capital-intensive and good Y as labor-intensive (hence abstracting from factor-intensity reversals). The analysis begins by considering the case of free trade in goods under factor immobility assumptions when such trade is related to differences in relative factor endowments among countries. The domestic transformation curve is represented by T_dT_d in Figure 1 tangent to the free trade terms of trade line MN at point P_d, while the foreign transformation curve is represented by T_fT_f, tangent to MN at point P_f (OP_d and OP_f differ because of the different relative factor endowments among countries). Tastes are assumed to be identical and homothetic (to abstract from scale effects in consumption), and represented by the common consumption vector OZ. In free trade equilibrium, the trade triangle is P_dDC_d in the home country and P_fFC_f in the foreign one, the two triangles being equal.

Suppose that the home country imposes an autarkic tariff on imports of good X, so that general equilibrium with the tariff occurs at point Q on T_dT_d and Q' on T_fT_f. The shift in the production point from P_d to Q in the home country, indicated by an increase in the output of the capital-intensive good and a decrease in the output of the labor-intensive good, implies a rise in the absolute (and relative) price and marginal product of capital and a fall in the absolute (and relative) price and marginal product of labor; while the movement from P_f to Q' on T_fT_f implies the opposite in the foreign country. Points Q and Q' cannot represent equilibrium, however, if from this position of self-sufficiency all impediments to the movement of both capital and labor are removed, since in the home country the marginal product of capital will be greater and the marginal product of labor less than the respective marginal products of these factors in the foreign country. Clearly there is a basic indeterminacy in the model; either capital can move into the home country or labor move out of the home country, or both, until marginal products and factor prices are equalized between countries. But this indeterminacy can be easily resolved by allowing for the mobility of only one factor, the procedure used by both Mundell and Nadel, and is immaterial to the factor-owners in any event, since regardless of which factor is assumed to move and whether the factor-owner is assumed to move with his exported factor or not, the effect of the stipulated fac-

* Visiting associate professor of economics, Stanford University. I wish to thank William Der, Brian Hindley, and Harry Johnson for their most useful comments.

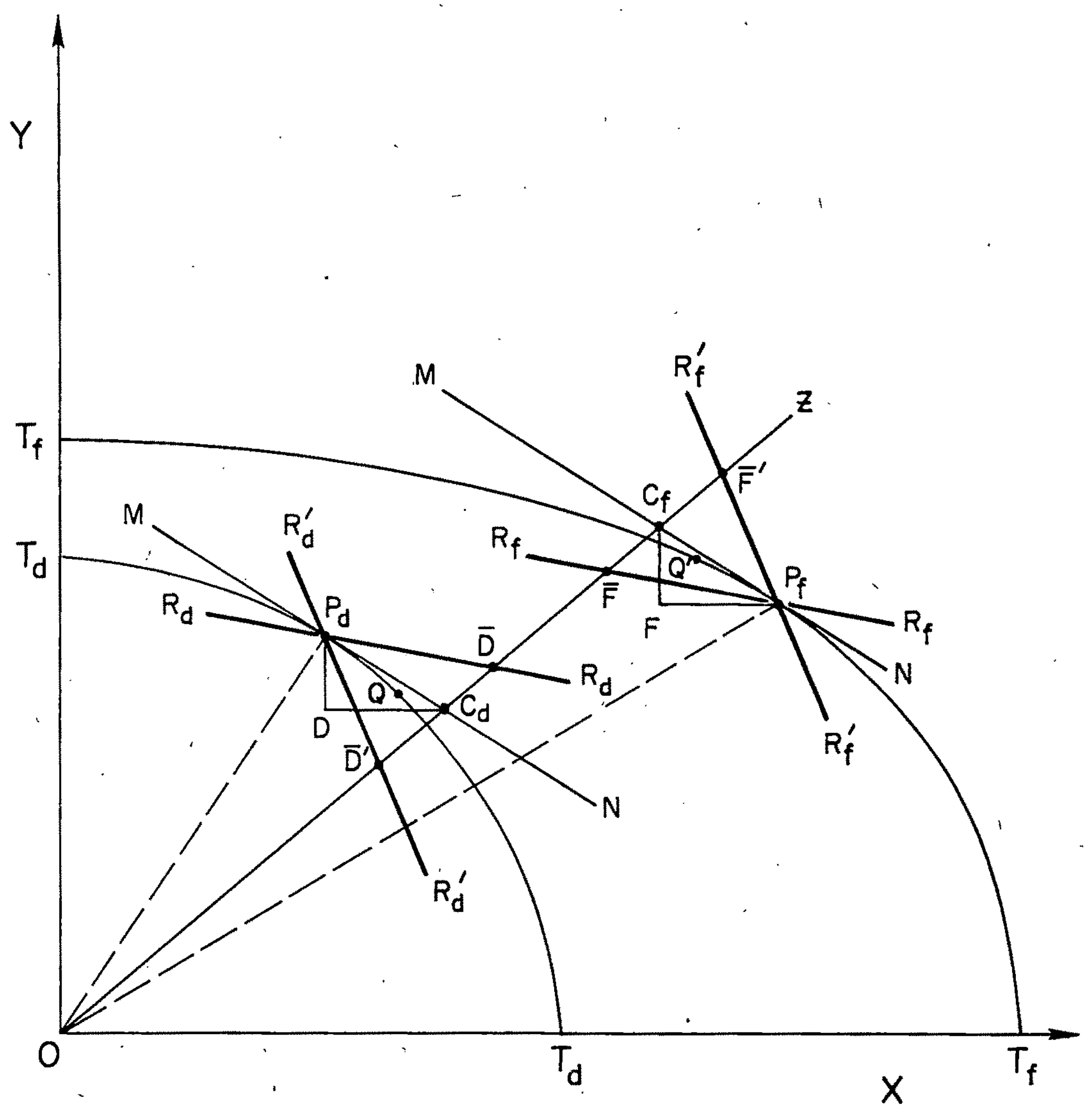


FIGURE 1

tor mobility on the economic well-being of all factor-owners will be the same, and equal to the effect that free trade has on them (it should be noted that in the case where factor-owners move the relevant criterion for establishing the equivalency of factor mobility and commodity trade must be the economic welfare of factor-owners not a country's aggregate consumption point).

The case where factor-owners move with their exported factors is considered in Figure 1. If only capital is assumed to be mobile, equilibrium in the home country will be determined where the Rybczynski line R_dR_d (defined as the output locus measured at constant prices as the absolute amount of one factor of production changes under *ceteris paribus* conditions) cuts the given consumption vector OZ at $\overline{D}$; while in the foreign country equilibrium is determined where the Rybczynski line R_fR_f cuts the OZ at $\overline{F}$ (because of the assumption of the identical technologies, R_dR_d is parallel to R_fR_f' and $P_d\overline{DF}P_f$ is a parallelogram; hence

$P_d\overline{D}=P_f\overline{F}$). Alternatively, if only labor is assumed mobile, domestic equilibrium occurs at $\overline{D}'$ where $R_d'R_d'$ cuts OZ; while foreign equilibrium occurs at $\overline{F}'$, where $R_f'R_f'$ cuts OZ. In both cases, equilibrium is characterized by (i) the equality of the slopes of the production vectors in the two countries, implying the equality of factor endowment ratios as well, and (ii) the equality of the slope of the common production vector with that of the common consumption vector on OZ; hence the tariff could be removed and there would be no incentive for commodity trade. It should be noted that it has been assumed above that *either* capital *or* labor moves in response to the factor price differentials existing between the two countries in closed-economy equilibrium; this, of course, is not necessary, and the same conclusion holds regardless of whether it is assumed that capital moves out and labor moves into the home country, or both factors move out, or both factors move in, so long as in final equilibrium the slopes of the production and consumption vectors are equal. Because of the basic indeterminacy in the model, we cannot tell what the precise pattern of intercountry factor mobility will be; but this is an irrelevancy for factor prices are equalized and factor mobility will substitute perfectly for

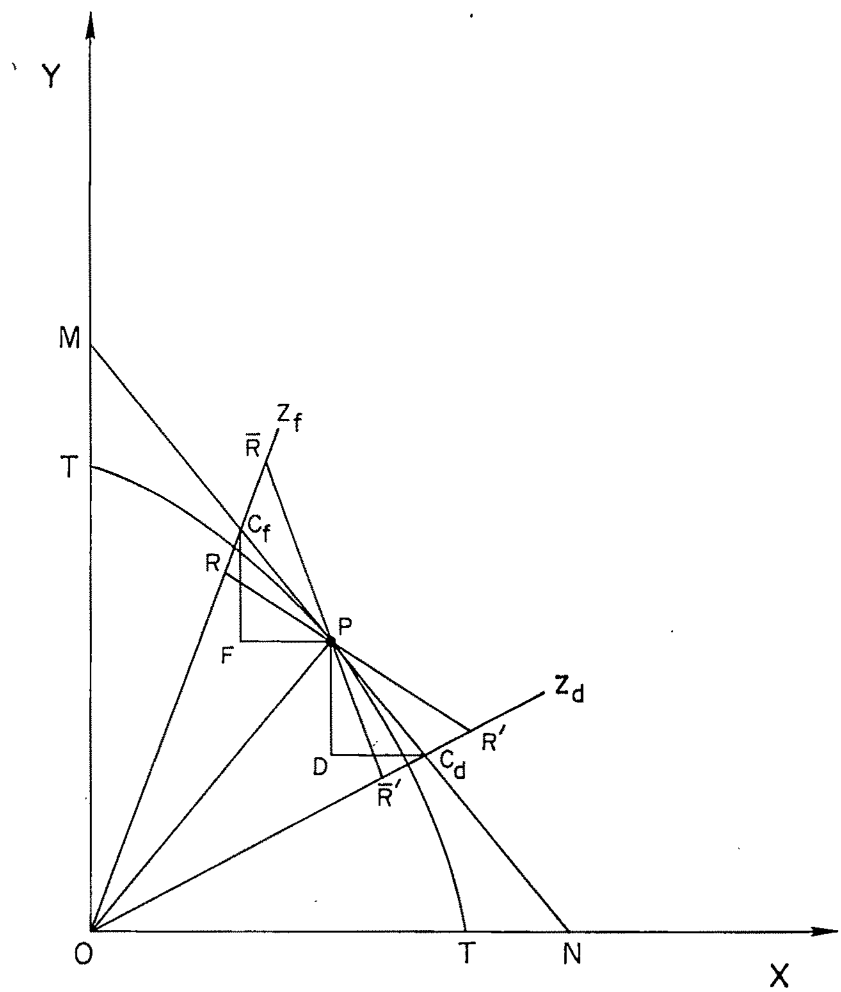


FIGURE 2

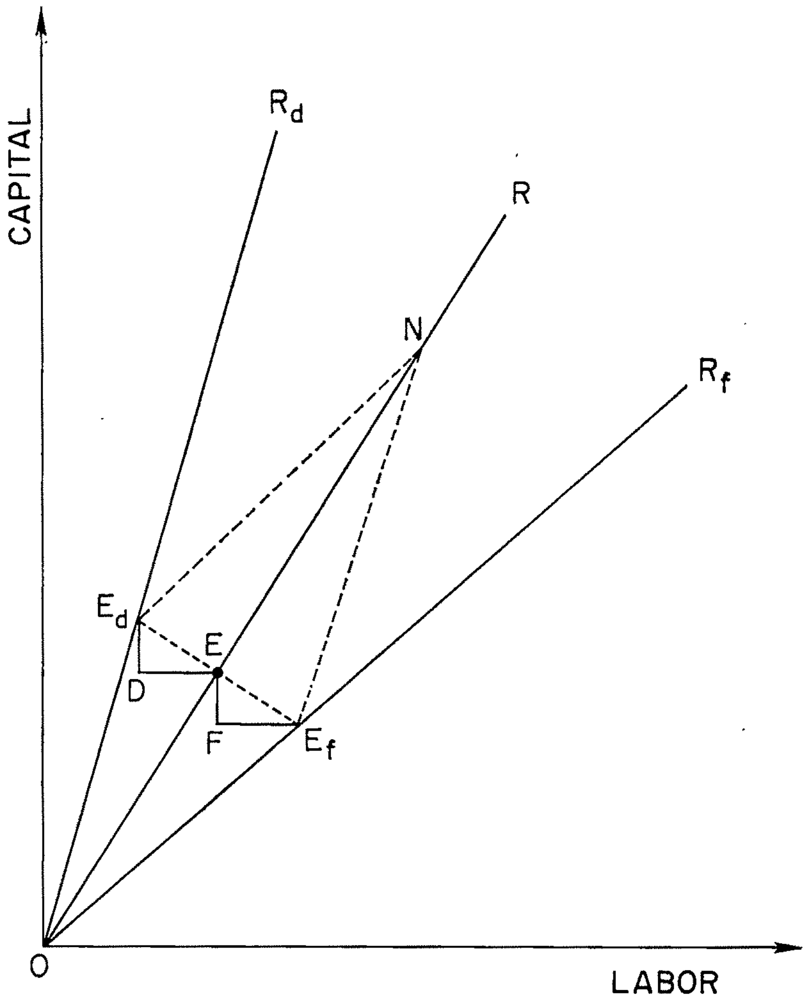


FIGURE 3

commodity trade with all possible patterns of factor movement consistent with the above equality.

The reader can satisfy himself of this fact by assuming an initial equilibrium with the slope of the common production vector equal to that of the common consumption vector. Factors can then be reallocated between the two countries. So long as the reallocations are proportionate in that they do not disturb the equal factor endowment ratios in the two countries, and thus the common production vectors, there will be no incentive for trade on the part of either country, and thus no commodity and factor price changes. Proportionate reallocations are the only relevant kind in this case, since regardless of the pattern of factor mobility equilibrium must fall on the vector OZ.

II

The interpretation of the Heckscher-Ohlin model that emphasizes relative factor endowment as *the* determinant rather than *a* determinant of international trade has come into disrepute as compared with the alternative interpretation that the model establishes in a disciplined manner a variety of potential

trade determinants, including differences in tastes between countries.[1] The case where trade is assumed to be due to differences in tastes is illustrated in Figure 2 where TT represents the common transformation curve in the two countries and P the common production point at the equilibrium terms of trade line MN which cuts the foreign consumption vector OZ_f at C_f and the domestic consumption vector OZ_d at C_d, determining the equilibrium (and thus equal) trade triangles C_fFP and PDC_d.

As before, assume that the home country imposes an autarkic tariff that raises the marginal product of capital in the home as compared with the foreign country and vice versa for the marginal product of labor. If from this position of self-sufficiency, the obstacles to factor mobility are removed, it can be shown that in this case there is only one pattern of factor flows consistent with factor price equalization at the given consumption vectors OZ_f and OZ_h (this assumes that factor-owners adopt the tastes of the country of immigration)—not an infinite number of mirror image solutions. For example, it is clear from the geometry of Figure 2 that the flow of capital alone, from the foreign to the home country, is inconsistent with equilibrium on OZ_f and OZ_d, since RP must be smaller than PR' given that $C_fP=PC_d$; and a similar proof can be employed to show the inconsistency of equilibrium on OZ_f and OZ_d when only labor is assumed to be mobile, since $\overline{R}P$ must be larger than $P\overline{R}'$.

The equilibrium pattern of factor flows is illustrated in Figure 3 where point E represents the initial common factor endowment point on OR, the initial common factor endowment ratio, and OR_d and OR_f the factor endowment ratios in the two countries that correspond to the production vectors OC_d and OC_f required for the mirror image solution. The world factor endowment point N is found by doubling OE, and the required post-mobility factor endowment points E_d and E_f determined by vector addition. Since E_dE_f is one diagonal of the parallelogram OE_dNE_f and ON the other, with $OE=EN$, the points E_dEE_f all must lie on a straight line such that $E_dE=EE_f$. The home country must "import" E_dD of capital and "export" DE of labor, while the factor "trade triangle" of the foreign country is EFE_f, with $EFE_f = E_dDE$.

[1] See my unpublished paper and also see James Melvin.

REFERENCES

F. Flatters, "Commodity Price Equalization: A Note on Factor Mobility and Trade," *Amer. Econ. Rev.*, June 1972, *62*, 473–76.

M. B. Krauss, "The Heckscher-Ohlin Model of International Trade: An Interpretation," mimeo., Stanford Univ. 1974.

J. R. Melvin, "Commodity Taxation as a Determinant of Trade," *Can. J. Econ.*, Feb. 1970, *3*, 62–78.

R. A. Mundell, "International Trade and Factor Mobility," *Amer. Econ. Rev.*, June 1957, *47*, 321–35.

E. Nadel, "International Trade and Capital Mobility," *Amer. Econ. Rev.*, June 1971, *61*, 368–79.

———, "Capital Goods, Intermediate Goods, and the Volume of Trade," *Can. J. Econ.*, May 1971, *4*, 238–49.

P. A. Samuelson, "International Trade and the Equalization of Factor Prices," *Econ. J.*, June 1948, *58*, 163–84.

Choice Involving Unwanted Risky Events and Optimal Insurance: Comment

By David S. Salkever*

In a recent paper in this *Review*, J. M. Parkin and S. Y. Wu (P-W) analyzed the demand for insurance in a model with two uncertain states of the world, A and B. They assumed that an unwanted event (say, illness) occurs in state B and they also permit the individual's conditional utility function to vary across these states. The purpose of this note is to show that while their mathematical statement of the problem was formally correct, their interpretation of results was erroneous. I shall briefly state a more accurate interpretation of their model's implications.

I. The P-W Model and Conclusions

P-W defined the individual's problem as the choice of the amount of insurance (denoted by the premium payment q) which would maximize expected utility E over states A and B where

$$(1)\quad E = pU(y-q) + (1-p)V(y'-q+I(q))$$

where p and $(1-p)$ are the probabilities associated with states A and B, respectively, U and V are the utility functions associated with these states, y and y' are the incomes associated with these states, and $I(q)$ is the insurance benefits payable if state B occurs.[1] Unconstrained maximization of E with respect to q yields the first-order condition:

$$(2)\quad pU_y + (1-p)V_{y'} = (1-p)I_qV_{y'}$$

where the subscripts denote derivatives.

P-W then proceed to examine the second-order conditions for a maximum of E and from this examination they draw the following conclusions.

> ... if both the utility of income functions are concave, the second-order condition is satisfied and insurance will be purchased. Also ...
> ... if both the utility of income functions are convex, the second-order condition is violated; then a solution of no-insurance becomes optimal. [p. 985]

II. Reinterpretation and Discussion

Both of these conclusions, however, are erroneous because in fact the satisfaction of P-W's second-order conditions implies nothing about the purchase of insurance as it is usually defined. More specifically, the purchase of insurance as usually defined means $q^*>0$ where q^* is the optimal amount of insurance. That is, it means that the individual parts with a certain positive amount of money in return for a contingent sum payable if the unwanted event occurs (i.e., in state B).[2]

To see that concavity need not imply $q^*>0$, let us consider the case analyzed by P-W where $I=zq/(1-p)$, and z is a constant such that $1-p<z<1$. For simplicity, we assume that U_y and $V_{y'}$ are linear functions denoted by $(a+by)$ and $(c+dy')$, respectively. In this case, the first-order condition (2) can be solved for q^* yielding

* Assistant professor, department of medical care and hospitals and department of political economy, Johns Hopkins University.

[1] U and V are indirect utility functions with price arguments suppressed since prices are not assumed to vary across states of the world.

[2] Of course, insurance is usually defined in this way because a market for negative insurance (contingent repayment loans) does not exist. This is probably due not to an absence of demand—consider the case of individuals in occupations whose returns are highly uncertain—but to a kind of market failure which arises from the relationship between information costs and incentives for transmitting information created by the insurance contract. The individual can obtain information on the actual state of the world at little cost but under negative insurance he has an incentive to conceal this information from the insurer. For the insurer, on the other hand, information about the actual state of the world is relatively costly. This implies that enforcement of negative insurance contracts would be difficult.

$$(3)\quad q^* = \frac{(z+p-1)(c+dy') - p(a+by)}{-(z+p-1)^2 d/(1-p) - pb}$$

Noting that concavity of U and V implies $b<0$ and $d<0$, and that $(a+by)$ and $(c+dy')$ are the conditional marginal utilities of income in states A and B, respectively, when no insurance is purchased, it is obvious that q^* will not in fact be positive unless the occurrence of the unwanted event causes a sufficiently large increase in the marginal utility of income in the absence of insurance.[3,4]

This conclusion is hardly surprising. The basic purpose of insurance is to redistribute income claims across states of the world. It will only be purchased if it redistributes these claims from states with low marginal utilities to states with high marginal utilities.[5]

Turning to P-W's second conclusion, it is easy to see that convexity of U and V does not imply that no insurance ($q^*=0$) is optimal; instead, it implies that an interior maximum does not exist and either $q^*=K$ or $q^*=0$ is optimal where K is a positive upper limit on insurance purchases which will presumably be related to the individual's wealth. The point is that a given incremental transfer of claims from the low-marginal-utility state to the high-marginal-utility state will increase the disparity in marginal utility between the two states. Each subsequent incremental transfer will add to E by an increasing amount and no equilibrium will be reached.[6]

In summary, the decision of whether or not to insure against an unwanted event depends upon the effect of that event on the marginal utility of income. The concavity or convexity of utility functions implies nothing in particular about this decision except that the decision may represent a corner solution. But even here knowledge of concavity or convexity need not imply that some positive purchase of insurance is optimal.

It is interesting to note the close analogy between our analysis and the literature on the option value question. In both cases, we are dealing with conditional utility functions that differ across states of the world and with a certain payment exchanged for uncertain benefits. In both cases it has been shown that concavity of the conditional utility functions is not sufficient to guarantee that the quantity in question (insurance purchases or option value) will be positive.[7]

[3] In the fixed-loading-charge case also considered by P-W, where $I=(q-k)/(1-p)$ and k is the loading charge, the solution of (2) yields

$$q^* = \frac{(c+dy) - (a+by) - dk/(1-p)}{-pd/(1-p) - b}$$

Here q^* could still be positive, with $d<0$ and $b<0$, even if $c+dy<a+by$. But (2) may only lead us to a local maximum; it indicates the value of q^* given that insurance is purchased and the charge k is incurred. The situation with no insurance and no loading charge may be preferable even if (2) indicates that q^* is positive.

[4] In fairness to P-W, it should be pointed out that their analysis suggests they were not totally unaware of this conclusion. For example, the assumption stated in their equation (5), p. 983, that income in the absence of insurance is lower in state B if combined with the assumption (not made by P-W) that $U_y=V_y$ for any given y would be sufficient to guarantee $q^*>0$ when insurance is actuarially fair and U and V are concave. However, since a conclusion contradictory to ours—that concavity is sufficient to guarantee that insurance will be purchased—is stated without qualification, we can at least say that their exposition is unclear on this point.

[5] And if loading charges are not too high.

[6] Assuming that marginal loading charges do not increase with the amount of insurance purchased.

[7] See Richard Schmalensee for a discussion of this result in the option value case.

REFERENCES

J. M. Parkin and S. Y. Wu, "Choice Involving Unwanted Risky Events and Optimal Insurance," *Amer. Econ. Rev.*, Dec. 1972, *62*, 982–87.

R. Schmalensee, "Option Demand and Consumer's Surplus: Valuing Price Changes Under Uncertainty," *Amer. Econ. Rev.*, Dec. 1972, *62*, 813–24.

Choice Involving Unwanted Risky Events and Optimal Insurance: Reply

By J. M. Parkin and S. Y. Wu*

In his comment David Salkever pointed out that we have incompletely interpreted our second-order conditions and stated that the concavity of the utility of income functions in both states need not imply a positive optimum insurance, and the convexity of these functions in both states does not imply absence of insurance. We agree with these criticisms and welcome his correction.

Assuming the marginal utility of income functions are linear, Salkever suggests that the consumer will purchase insurance if, and only if,

(1) (i) For $I(q) = \dfrac{q}{1-\pi}$,

$$\left.\frac{\partial U(M)}{\partial q}\right|_{M=y} < \left.\frac{\partial V(M')}{\partial q}\right|_{M'=y'}$$

(ii) For $I(q) = \dfrac{\beta}{1-\pi}q$,

$$\left.\frac{\partial U(M)}{\partial q}\right|_{M=y} < \left(\frac{\beta+\pi-1}{\pi}\right)\left.\frac{\partial V(M')}{\partial q}\right|_{M'=y'}$$

* Professors of economics, University of Manchester, England, and University of Iowa, respectively.

With the marginal utility of income nonlinear, however, condition (1) is still not sufficient. A global optimum condition for insurance should be

(2) (i) For $I(q) = \dfrac{q}{1-\pi}$,

$$\left.\frac{\partial U(M)}{\partial q}\right|_{M=y-q} < \left.\frac{\partial V(M')}{\partial q}\right|_{M'=y'-q+I(q)}$$

for some M and M'

(ii) For $I(q) = \dfrac{\beta}{1-\pi}q$,

$$\left.\frac{\partial U(M)}{\partial q}\right|_{M=y-q} < \left(\frac{\beta+\pi-1}{\pi}\right)\left.\frac{\partial V(M')}{\partial q}\right|_{M'=y'-q+I(q)}$$

for some M and M'

The interpretation of equation (2) is that although the consumer may find it nonoptimal to purchase a small amount of insurance, it is advantageous for him to purchase an amount exceeding a certain critical sum up to the limit $q^*=y$.

As Salkever rightly points out, both conditions (1) and (2) do not depend on the second derivatives of the utility functions.

The Number of Firms and Competition: Comment

By HAJIME HORI*

In an article in this *Review*, Eugene Fama and Arthur Laffer assert that "when there are two or more noncolluding firms in the industry, . . . under certain conditions a general equilibrium in fact implies . . . that the industry equilibrium is competitive" (p. 671). By competition is meant a situation in which "[an] individual firm's output decision literally has no effect on the market price of the good" (p. 670). This will be a startling result if it is true.

Their reasoning is as follows: If a firm in a certain industry increases its output from an original equilibrium level, the price of the good produced by this industry will fall. Then, "returns at least to some factors would . . . be lower in this industry than in others" (p. 673). If the production function is linearly homogeneous, factor returns in this industry will be equal to those in other industries if, and only if, other firms in this industry contract their output by exactly that amount by which our first firm has increased its output. Thus mobility of factors insures that the total output of this industry will come back to its original level, and therefore that the price of its output cannot be affected by an individual firm.

An apparent slip in this reasoning lies in the direct association of the product price or the total revenue with "factor rewards in this industry." Namely, it is implicitly assumed that the revenues of all the firms in this industry are all imputed to and distributed among the factors. Since all the factors are assumed to be mobile among industries, this amounts to assuming that the firms cannot retain any residual profits. Does the model of Fama and Laffer have any special structure that justifies this procedure?

Their certain conditions relate to divisibility and mobility of factors of production, maximization of returns by factors, stability of equilibrium points, identical production technology among all the firms in a given industry, and linear homogeneity of this production technology, and there is nothing special about these conditions that guarantees the distribution of all the revenues among mobile factors. Also, although they emphasize the importance of their general equilibrium approach, it simply means that a firm under consideration is assumed to anticipate the reactions of other firms in the same industry when it decides its own output level. Since their analysis is supposed ". . . not to depend on any specific adjustment process by which other firms respond to output changes of an individual firm" (pp. 673–74), their approach does not exclude the classical Stackelberg duopoly solution, in which both the leader and the follower usually enjoy some extra profits.

It is clear now that the essential assumption of Fama and Laffer which has led them to their main conclusion is that the firms cannot retain any residual profits. The total output of the industry and its price are determined not by the profit maximizing behavior of the firms but by the condition that the firms have zero profits. It is difficult to accept this model as a picture of the capitalistic mode of production and competition.

* Assistant professor, Tulane University. I have benefited from discussions with George Borts and comments by an anonymous referee.

REFERENCE

E. F. Fama and A. B. Laffer, "The Number of Firms and Competition," *Amer. Econ. Rev.*, Sept. 1972, *62*, 670–74.

The Number of Firms and Competition: Reply

By EUGENE F. FAMA AND ARTHUR B. LAFFER*

Hajime Hori's main criticism of our paper appears in the third paragraph of his comment.

> An apparent slip in this reasoning lies in the direct association of the product price or the total revenue with "factor rewards in this industry." Namely, it is implicitly assumed that the revenues of all the firms in this industry are all imputed to and distributed among the factors. Since all the factors are assumed to be mobile among industries, this amounts to assuming that the firms cannot retain any residual profits. [p. 805]

We believe that in any well-specified model all of the firm's revenues are imputed to factors. There may be a factor, typically called ownership, whose claims against the firm's revenues are of a residual nature, and the return to this factor may be called "profit." The implication of our analysis is that in a general equilibrium, profit per unit of this factor must be the same across all firms and industries.

It is not the case, however, that profits must be distributed. The managers of the firm, on behalf of the owners, may decide that part of the profits are to be reinvested in the firm. In this case the reinvested profits are treated as new resources provided by the owners of the firm. But the basic point still holds: total, that is, preretention profit per unit of ownership must be the same across all firms and industries. This is the general equilibrium implication of our assumption that all factors are completely mobile across firms and industries.

In the second to last paragraph of his paper, Hori injects another criticism of our work. "Since their analysis is supposed 'not to depend on any specific adjustment process by which other firms respond to output changes of an individual firm,' their approach does not exclude the classical Stackelberg duopoly solution, in which both the leader and the follower usually enjoy some extra profits" (p. 805).

In our world, if the leader and follower enjoy extra profits, then these extra profits must be the same for all other firms and industries. Otherwise, there will be entry. The Stackelberg duopoly model restricts entry. As we state in at least two places, our model does not. Perhaps the absence of a restriction on entry makes our results trivial. But Hori's comments do not convince us that our results are, as he claims, incorrect.

* Graduate School of Business, University of Chicago.

NOTES

EIGHTY-SEVENTH ANNUAL MEETING OF THE AMERICAN ECONOMIC ASSOCIATION

San Francisco, California, December 27–30, 1974

Preliminary Announcement of the Program

Friday, December 27, 1974

10:00 A.M. EXECUTIVE COMMITTEE MEETING

Saturday, December 28, 1974

8:30 A.M. DIVERSIFICATION OF COMMERCIAL BANK AND NONBANKING ACTIVITIES* (Joint Session with the American Finance Association)
Chairperson: SHERMAN J. MAISEL, University of California, Berkeley, and National Bureau of Economic Research
Papers: SAMUEL CHASE AND JOHN MINGO, Federal Reserve Board
The Regulation of Bank Holding Companies
THOMAS I. STORRS, North Carolina National Bank
More Freedom for Banks
Discussants: ANDREW F. BRIMMER, Harvard University
DANIEL H. BRILL, Commercial Credit Corporation
JOAN WALTERS, Fairfield University

8:30 A.M. DOMESTIC ADJUSTMENT TO BALANCE-OF-PAYMENTS DISEQUILIBRIUM (Joint Session with the American Finance Association)
Chairperson: CHARLES P. KINDLEBERGER, Massachusetts Institute of Technology
Paper: MARINA v. N. WHITMAN, University of Pittsburgh
Domestic Adjustment to Balance-of-Payments Disequilibrium: What Have We Learned?
Discussants: RUDIGER DORNBUSCH, University of Chicago
WILLIAM J. FELLNER, Council of Economic Advisers
ALEXANDER KAFKA, International Monetary Fund and University of Virginia
PAUL A. VOLCKER, U.S. Treasury

8:30 A.M. POLITICAL ECONOMY OF SUBSIDIZED CREDITS IN EAST-WEST TRADE* (Joint Session with the Association for the Study of the Grants Economy)
Chairperson: PAUL MARER, Indiana University
Papers: JANOS HORVATH, Butler University
Are Eximbank Credits Subsidized?—Toward an Empirical Analysis
DOUGLAS R. BOHI, Southern Illinois University
Evaluation of the Role of Eximbank in Promoting Exports
THOMAS A. WOLF, Ohio State University
East-West Trade Credit Policy: A Comparative Analysis
(Author to be announced)
U.S. National Interest and Credits in East-West Trade
Discussants: DAVID G. WIGG, Export-Import Bank
LAWRENCE J. BRAINARD, Chase Manhattan Bank
WILSON E. SCHMIDT, Virginia Polytechnic Institute

8:30 A.M. THE ECONOMICS OF LABOR: AN ASSESSMENT OF RECENT RESEARCH (Joint Session with the Industrial Relations Research Association)
Chairperson: ORLEY C. ASHENFELTER, Princeton University
Papers: GLEN G. CAIN, University of Wisconsin
The Dual and Radical Challenge to Labor Economics
GEORGE E. JOHNSON, University of Michigan
The Economics of Trade Unionism

* Not to appear in the *Papers and Proceedings*

H. GREGG LEWIS, University of Chicago
The Economics of Time and Labor Supply
Discussants: DAVID M. GORDON, New School for Social Research
GEORGE H. HILDEBRAND, Cornell University
PAULA STEPHAN, Georgia State University

10:30 A.M. CHINA AND INDIA: COMPARATIVE DEVELOPMENT DURING THE LAST TWENTY-FIVE YEARS
Chairperson: WALTER P. FALCON, Food Research Institute, Stanford University
Papers: BARRY M. RICHMAN, University of California, Los Angeles
Chinese and Indian Development: An Interdisciplinary Environmental Analysis
THOMAS E. WEISSKOPF, University of Michigan
China and India: Contrasting Experiences in Economic Development
Discussants: PADMA DESAI, Russian Research Center, Harvard University
ROBERT L. HEILBRONER, New School for Social Research
JOHN G. GURLEY, Stanford University

10:30 A.M. TWENTY-FIVE YEARS AFTER THE REDISCOVERY OF MONEY: WHAT HAVE WE LEARNED? (Joint Session with the American Finance Association and the Econometric Society)
Chairperson: WILLIAM POOLE, Brown University
Papers: STANLEY FISCHER, Massachusetts Institute of Technology
Developments in Monetary Theory
WILLIAM C. BRAINARD AND RICHARD N. COOPER, Yale University
Views on the Effectiveness of Monetary Policy: Then and Now
Discussants: MILTON FRIEDMAN, University of Chicago
FRANCO MODIGLIANI, Massachusetts Institute of Technology

10:30 A.M. INFLATION IN COMPARATIVE PERSPECTIVE* (Joint Session with the Association for Comparative Economic Studies)
Chairperson: MORRIS BORNSTEIN, University of Michigan
Papers: EDWARD AMES, State University of New York, Stony Brook
The Economics of Open vs. Repressed Inflation
JUDITH THORNTON, University of Washington
Inflation in a Centrally Planned Economy
CARLOS DIAZ-ALEJANDRO, Yale University
Inflation in Less Developed Market Economies
Discussants: DONALD R. HODGMAN, University of Illinois
GEORGE GARVY, Federal Reserve Bank of New York

10:30 A.M. OPTIMIZING MODELS OF PUBLIC DECISION MAKING (Joint Session with the Public Choice Society)
Chairperson: ROGER G. NOLL, CALIFORNIA INSTITUTE OF TECHNOLOGY
Papers: JOHN FEREJOHN AND MORRIS FIORINA, California Institute of Technology
Optimizing Models of Congress
MARK ROBERTS, Harvard University
The Behavior of Public Organizations
Discussants: OTTO A. DAVIS, Carnegie-Mellon University
PAUL JOSKOW, Massachusetts Institute of Technology

10:30 A.M. THE NATION'S HOUSING NEEDS AND HOUSING DEMAND INTO THE 1980's* (Joint Session with the American Real Estate and Urban Economics Association)
Chairperson: FRANK S. KRISTOF, New York State Urban Development Corporation
Papers: BERNARD FRIEDEN, Joint Center for Urban Studies, Massachusetts Institute of Technology
Forecasting the Nation's Housing Needs through 1980
THOMAS C. MARCIN, Forest Service, U.S. Department of Agriculture, AND JOHN KOKUS, JR., American University
Economic and Demographic Factors in Housing Needs and Demand into the 1980's
Discussants: WALLACE F. SMITH, University of California, Berkeley
GRACE MILGRAM, New York State Urban Development Corporation

12:30 P.M. JOINT LUNCHEON (With the American Finance Association)
Chairperson: JOHN LINTNER, Harvard University
Speaker: H. JOHANNES WITTEVEEN, Managing Director, International Monetary Fund

* Not to appear in the *Papers and Proceedings*

2:00 P.M. THE HUMAN CAPITAL APPROACH: AN APPRAISAL (Joint Session with the Econometric Society)
Chairperson: CLAIR VICKERY, University of California, Berkeley
Papers: FINIS WELCH, University of California, Los Angeles, and the RAND Corporation
Theory of Human Capital: Life Cycles, Education, and Discrimination
SAMUEL BOWLES AND HERBERT GINTIS, University of Massachusetts
The Problem with Human Capital Theory
Discussants: RONALD OAXACA, University of Massachusetts
LOURDES BENERIA SURKIN, Columbia University
MARTIN CARNOY, Stanford University and The Center for Economic Studies

2:00 P.M. TAXATION OF NATURAL RESOURCES (Joint Session with the American Finance Association)
Chairperson: JOSEPH A. PECHMAN, The Brookings Institution
Papers: CHARLES E. MCLURE, JR., Rice University
The Incidence of World Taxes on Natural Resources
GERARD BRANNON, JR., Georgetown University
U.S. Taxes on Energy Resources
Discussants: ARNOLD C. HARBERGER, University of Chicago
GLEN P. JENKINS, Harvard University
ARTHUR W. WRIGHT, University of Massachusetts

2:00 P.M. UNDERSTANDING WORLD INFLATION
Chairperson: ARTHUR M. OKUN, The Brookings Institution
Panelists: WILLIAM H. BRANSON, Princeton University
GEORGE L. PERRY, The Brookings Institution
RICHARD T. SELDEN, University of Virginia
RONALD L. TEIGEN, University of Michigan

2:30 P.M. EMERGING STRUCTURAL CHANGES IN TRANSPORT AND PUBLIC UTILITIES (Joint Session with the Transportation and Public Utilities Group)
Chairperson: L. L. WATERS, Indiana University
Papers: PAUL W. CHERINGTON, Harvard University
Implications of the Railroad Reorganization Act
LEON N. MOSES AND G. S. GOLDSTEIN, Northwestern University
Transportation Controls and the Spatial Structure of Urban Areas
PAUL W. MACAVOY, Massachusetts Institute of Technology
Shifts in the Electric Power Industry
Discussant: ALFRED E. KAHN, Cornell University

2:30 P.M. INTERNATIONALIZATION OF COMMERCIAL BANKING* (Joint Session with the American Finance Association)
Chairperson: LAWRENCE B. KRAUSE, The Brookings Institution
Papers: ANDREW F. BRIMMER, Harvard University AND FREDERICK R. DAHL, Federal Reserve Board
Growth of American International Banking: Implications for Public Policy
ROBERT Z. ALIBER, University of Chicago
Monetary Interdependence Under Floating Exchange Rate Systems
JOHN R. HEWSON AND EISUKE SAKAKIBARA, International Monetary Fund
A General Equilibrium Analysis of the Eurocurrency Markets and Related Policy Issues
Discussants: ROBERT K. WILMOUTH, First National Bank of Chicago
EUGENE A. BIRNBAUM, Chase Manhattan Bank

8:00 P.M. RICHARD T. ELY LECTURE
Chairperson: R. A. GORDON, University of California, Berkeley
Speaker: ALICE M. RIVLIN, The Brookings Institution

Sunday, December 29, 1974

8:30 A.M. LAW AND ECONOMICS*
Chairperson: SIMON ROTTENBERG, University of Massachusetts
Papers: ARTHUR LEFF, Yale Law School
Economic Analysis of Law

* Not to appear in the *Papers and Proceedings*

RICHARD A. POSNER, University of Chicago Law School
The Use of Economics to Explain and Improve the Legal System
ROSS D. ECKERT, University of Southern California
Economics and the Law of the Sea
SAM PELTZMAN, University of Chicago
The Effects of Auto Safety Regulation

8:30 A.M. SOURCES AND USES OF PANELS OF MICRODATA (Joint Session with the Econometric Society)
Chairperson: GUY H. ORCUTT, Yale University and The Urban Institute
Papers: HERBERT S. PARNES, Ohio State University
The National Longitudinal Surveys: New Vistas for Labor Market Research
JAMES N. MORGAN, University of Michigan
Getting and Using Survey Data, Particularly Those from the University of Michigan's Survey Research Center
SHIRLEY KALLEK, Bureau of the Census
Potential Applications of Census Bureau Economic Series for Microdata Analysis
HAROLD W. WATTS, University of Wisconsin
The Experimental Panel Data Resources at the Poverty Institute Data Center
Discussants: BARBARA R. BERGMANN, University of Maryland
BENJAMIN A. OKNER, The Brookings Institution
RICHARD R. NELSON, Yale University

8:30 A.M. THE BEHAVIOR OF FIRMS IN OLIGOPOLISTIC MARKETS: IN THEORY AND PRACTICE
Chairperson: OLIVER E. WILLIAMSON, University of Pennsylvania
Papers: PAUL JOSKOW, Massachusetts Institute of Technology
Firm Decision-Making Processes and Oligopoly Behavior
MARTIN SHUBIK, Yale University
Oligopoly, Information, and Disequilibrium
Discussants: ROGER SHERMAN, University of Virginia
LESTER G. TELSER, University of Chicago
F. M. SCHERER, Federal Trade Commission

8:30 A.M. POWER AND JUSTICE IN GRANTING PROCESSES* (Joint Session with the Association for the Study of the Grants Economy)
Chairperson: KENNETH E. BOULDING, University of Colorado
Papers: ELIZABETH M. CLAYTON, University of Missouri-St. Louis
Property Rights Granted Under Alternative Economic Systems: A Taxonomy
WARREN J. SAMUELS, Michigan State University
Grants and the Theory of Power
DANIEL R. FUSFELD, University of Michigan
Immiserating the Poor
HOWARD TUCKMAN, Florida State University
The Economics of the Rich
Discussants: SAMUEL M. LOESCHER, Indiana University
MARTIN PFAFF, Wayne State University and the University of Augsburg

10:30 A.M. A REASSESSMENT OF DEVELOPMENT ECONOMICS
Chairperson: W. ARTHUR LEWIS, Princeton University
Panelists: IRMA ADELMAN, University of Maryland
HOLLIS B. CHENERY, International Bank for Reconstruction and Development
ANDREAS G. PAPANDREOU, York University
STEPHEN A. RESNICK, University of Massachusetts

10:30 A.M. ECONOMETRICS AND LABOR MARKET ANALYSIS* (Joint Session with the Industrial Relations Research Association)
Chairperson: SARA BEHMAN, California Polytechnic State University, San Luis Obispo
Panelists: ORLEY C. ASHENFELTER, Princeton University
BARBARA R. BERGMANN, University of Maryland
SAR A. LEVITAN, George Washington University
DANIEL J. B. MITCHELL, University of California, Los Angeles

* Not to appear in the *Papers and Proceedings*

10:30 A.M. A CRITICAL LOOK AT THE KEYNESIAN MODEL: THEORY AND APPLICATION
Chairperson: EVSEY D. DOMAR, Massachusetts Institute of Technology
Papers: ROBERT W. CLOWER AND AXEL S. B. LEIJONHUFVUD, University of California, Los Angeles
The Coordination of Economic Activities: A Keynesian Perspective
ROBERT EISNER, Northwestern University
The Keynesian Revolution Reconsidered
JAMES TOBIN, Yale University
The Legacies of Keynes
Discussants: PAUL DAVIDSON, Rutgers State University
ROBERT J. GORDON, Northwestern University
HYMAN P. MINSKY, Washington University

10:30 A.M. SUPPLY AND MOBILITY OF WOMEN ECONOMISTS
Chairperson: WALTER P. ADAMS, Michigan State University
Papers: COLLETTE MOSER, Michigan State University AND ALICE H. AMSDEN, University of California, Los Angeles
Job Search and Affirmative Action
MYRA H. STROBER, Stanford University
Women Economists: Education, Training, and Early Career Aspirations
BARBARA B. REAGAN, Southern Methodist University
Career Patterns of Women Economists
BARBARA B. REAGAN, Southern Methodist University
Report on Activities of the Committee on the Status of Women in the Economics Profession

10:30 A.M. GOVERNMENT REGULATION AND PRICE SETTING IN THE HEALTH CARE SECTORS* (Joint Session with the Health Economics Research Organization)
Chairperson: ROBERT L. ROBERTSON, JR., Mount Holyoke College
Papers: J. JOEL MAY, University of Chicago
The Impact of Regulation on the Hospital Industry
PAUL B. GINSBURG, Michigan State University
Price Controls and Hospital Costs
DONALD E. YETT, University of Southern California; JOHN M. MARSHALL, University of California, Santa Barbara, and University of Southern California; AND JOHN S. GREENLEES, University of Southern California
Setting Nursing Home Reimbursement Rates Under Medicare and Medicaid
Discussants: BARRY R. CHISWICK, Council of Economic Advisers and Queens College, City University of New York
GERALD ROSENTHAL, Bureau of Health Services Research, U.S. Department of Health, Education, and Welfare
KENNETH M. MCCAFFREE, Battelle Memorial Institute and University of Washington

10:30 A.M. COMPARISON OF FINANCIAL SYSTEMS* (Joint Session with the American Finance Association and the Economic History Association)
Chairperson: RONDO CAMERON, Emory University
Papers: RAYMOND W. GOLDSMITH, Yale University
The Comparative Quantitative Study of Financial Systems
ROGER H. HINDERLITER AND HUGH ROCKOFF, Rutgers University
The Investment Policies of Banks in the Leading Industrial Nations, 1888–1908
HOUSTON H. STOKES AND HUGH NEUBURGER, University of Illinois at Chicago Circle
German Banking in a Comparative Context: Costs and Benefits of the Kredit Bank in the Pre-1914 Era
Discussant: RICHARD E. SYLLA, North Carolina State University, Raleigh

12:30 P.M. JOINT LUNCHEON HONORING NOBEL PRIZEWINNER WASSILY W. LEONTIEF (With the Econometric Society)
Chairperson: JOHN KENNETH GALBRAITH, Harvard University
Introduction: ANNE P. CARTER, Brandeis University
Response: WASSILY W. LEONTIEF, Harvard University

* Not to appear in the *Papers and Proceedings*

2:00 P.M. THE PRINCIPLES COURSE: WHAT SHOULD BE IN IT AND WHERE SHOULD IT BE GOING?
Chairperson: G. L. BACH, Stanford University
Panelists: KENNETH E. BOULDING, University of Colorado
JOHN KENNETH GALBRAITH, Harvard University
JOHN G. GURLEY, Stanford University
W. LEE HANSEN, University of Wisconsin

2:00 P.M. WHAT HAVE WE LEARNED FROM ECONOMIC GROWTH THEORY? (Joint Session with the Econometric Society)
Chairperson: JOHN W. KENDRICK, George Washington University and National Bureau of Economic Research
Papers: RONALD FINDLAY, Columbia University
Implication of Economic Growth Theory for Trade and Development
DONALD J. HARRIS, University of Wisconsin
The Theory of Economic Growth: A Critical Evaluation
RICHARD R. NELSON, Yale University
Growth Theory from an Evolutionary Perspective
Discussants: ROBERT L. HEILBRONER, New School for Social Research
DALE W. JORGENSON, Harvard University

2:00 P.M. SOCIALISM IN LESS DEVELOPED COUNTRIES* (Joint Session with the Association for Comparative Economic Studies)
Chairperson: (To be announced)
Papers: PADMA DESAI, Harvard University AND JAGDISH BHAGWATI, Massachusetts Institute of Technology
Socialism in LDCs: The Case of India
EDMUND CLARK, Harvard University
Socialism in LDCs: The Case of Tanzania
BENT HANSEN, University of California, Berkeley
Socialism in LDCs: The Case of Egypt
Discussants: ALBERT O. HIRSCHMAN, Harvard and Princeton Universities
CARLOS DIAZ-ALEJANDRO, Yale University

2:00 P.M. CURRENT PROBLEMS IN PUBLIC FINANCE* (Joint Session with the American Finance Association)
Chairperson: NANCY TEETERS, Library of Congress
Papers: UWE E. REINHARDT, Princeton University
Alternate Benefit Programs for Health Insurance with Special Reference to the Canadian and German Experience
BENJAMIN A. OKNER, The Brookings Institution
Reform of the Payroll Tax
RICHARD NATHAN, The Brookings Institution
State and Local Government Use of General Revenue-Sharing Funds
Discussants: DOROTHY P. RICE, Social Security Administration
GAIL R. WILENSKY, University of Michigan
WAYNE VROMAN, University of Maryland
LLOYD ULLMAN, University of California, Berkeley

2:00 P.M. PAST DEVELOPMENTS AND FUTURE PROSPECTS FOR ETHNIC MINORITY GROUPS (Joint Session with the Caucus of Black Economists)
Chairperson: KARL D. GREGORY, Oakland University
Papers: ROBERT W. FOGEL, University of Chicago and University of Rochester
Taking Stock of Slavery and Its Aftermath
BERNARD E. ANDERSON, University of Pennsylvania AND PHYLLIS A. WALLACE, Massachusetts Institute of Technology
Public Policy and Black Economic Progress: Review of the Evidence
DAVID SWINTON, State University of New York, Stony Brook
Factors Affecting Future Prospects for Minority Groups
Discussants: RONALD L. TROSPER, University of Washington
DANIEL R. FUSFELD, University of Michigan
THOMAS SOWELL, University of California, Los Angeles

* Not to appear in the *Papers and Proceedings*

4:00 P.M. BUSINESS MEETING

8:00 P.M. PRESIDENTIAL ADDRESS
Chairperson: KERMIT GORDON, The Brookings Institution
Speaker: WALTER W. HELLER, University of Minnesota

Monday, December 30, 1974

8:30 A.M. PERSONAL SAVINGS IN THE POSTWAR WORLD: IMPLICATIONS FOR THE THEORY OF HOUSEHOLD BEHAVIOR
Chairperson: SUSAN W. BURCH, Federal Reserve Board
Papers: F. THOMAS JUSTER, University of Michigan AND LESTER D. TAYLOR, University of Arizona
The Role of Price Expectations in U.S. Postwar Saving Behavior
BURKHARD STRUMPEL, University of Michigan
Postwar Saving Behavior in the Western World
MICHAEL R. DARBY, University of California, Los Angeles
U.S. Saving During the Postwar Period
Discussants: THOMAS MAYER, University of California, Davis
(Others to be announced)

8:30 A.M. BIOGRAPHY AND AUTOBIOGRAPHY IN THE STUDY OF INTELLECTUAL HISTORY* (Joint Session with the History of Economics Society)
Chairperson: THOMAS SOWELL, University of California, Los Angeles
Papers: WILLIAM JAFFE, York University
Biography: A Genetic Component of Economic Analysis—Answer to a Challenge
THEODORE W. SCHULTZ, University of Chicago
Reflections on Lingering Doubts About Economics
FRITZ MACHLUP, New York University
Reminiscences Without Regrets
Discussant: WILLIAM BREIT, University of Virginia

8:30 A.M. GRANTS AND EXCHANGES IN HEALTH CARE PROGRAMS* (Joint Session with the Association for the Study of the Grants Economy and the Health Economics Research Organization)
Chairperson: TIBOR SCITOVSKY, Stanford University
Papers: KAREN P. DAVIS, Harvard University and The Brookings Institution
Distributional Implications of Medical Care Programs: The Case of Medicare and Medicaid
PHILIP JACOBS, Sir George Williams University
The Philanthropic Firm and the Economics of Giving
WILLIAM B. NEENAN, University of Michigan
On the Perceived Utilities and Disutilities of Government
THEODORE TSUKAHARA, JR., Pomona College AND WILLIAM WADMAN, Claremont Graduate School
Towards a Theory of Quality with Special Reference to Health
Discussants: BRUCE STUART, University of Massachusetts
DAVID C. WARNER, Yale University
THOMAS R. IRELAND, University of Missouri, St. Louis
SYLVESTER E. BERKI, University of Michigan

8:30 A.M. EAST-WEST TECHNOLOGICAL TRANSFER ISSUES AND PROBLEMS (Joint Session with the Association for Comparative Economic Studies)
Chairperson: ALEXANDER ECKSTEIN, University of Michigan
Papers: EDWIN MANSFIELD, University of Pennsylvania
International Technology Transfer: Problems, Costs, and Policies
ED HEWETT, University of Texas
The Economics of Transfering Technology from Market to Centrally Planned Economies
THOMAS E. RAWSKI, University of Toronto
Problems of Technology Absorption in Chinese Industry
Discussants: BARBARA G. KATZ, State University of New York, Stony Brook
PETER HELLER, University of Michigan

* Not to appear in the *Papers and Proceedings*

10:30 A.M. WOMEN AND THE ECONOMY*
Chairperson: MYRA H. STROBER, Stanford University

Papers: MARNIE W. MUELLER, Wesleyan University
Economic Determinants of Female Participation in Volunteer Work
LUCY B. MALLAN, Social Security Administration
Experience and Earnings of Women Labor Force Entrants, 1961–1971
ROBIN L. BARTLETT, Denison University
The Impact of Labor Force Participation on the Distribution of Family Earnings, 1962–1971
Discussants: FRANCINE D. BLAU, Ohio State University
HEIDI H. COCHRAN, New School for Social Research
BURTON A. WEISBROD, University of Wisconsin

10:30 A.M. MICROECONOMIC THEORY: CONFLICT AND CONTRACT
Chairperson: SIDNEY G. WINTER, JR., University of Michigan
Papers: JAMES M. BUCHANAN, Virginia Polytechnic Institute and State University
A Contractarian Paradigm for Economic Theory
DUNCAN K. FOLEY, Stanford University
"Problems" or "Conflicts?"
ALVIN K. KLEVORICK, Yale University
Law and Economic Theory: An Economist's View
Discussants: THEODORE C. BERGSTROM, Washington University, St. Louis
JOHN P. BROWN, Cornell University
NANCY L. SCHWARTZ, Northwestern University

10:30 A.M. HUMAN BEHAVIOR AND SOCIAL INSTITUTIONS: A RESEARCH PROGRAM OF THE NATIONAL BUREAU*
Chairperson: VICTOR R. FUCHS, Stanford University
Panelists: GARY S. BECKER, University of Chicago
WILLIAM M. LANDES, University of Chicago
ROBERT T. MICHAEL, Stanford University
JACOB MINCER, Columbia University
ROBERT WILLIS, National Bureau of Economic Research, Inc.

10:30 A.M. RECENT RESOURCE PROBLEMS OF LESS DEVELOPED COUNTRIES* (Joint Session with the American Agricultural Economics Association)
Chairperson: RONALD G. RIDKER, Resources for the Future
Papers: CHARLES PETER TIMMER, Food Research Institute, Stanford University
Energy and Food Supplies in LDCs
ERNEST STERN AND WOUTER TIMS, International Bank for Reconstruction and Development
Changes in the Relative Bargaining Strength of the LDCs
T. N. SHANKAR, Planning Commission, Government of India
Domestic Adjustments and Accommodations to Higher Raw Material and Energy Prices
Discussants: JOHN SCHNITTKER, Schnittker's Associates
RAYMOND F. MIKESELL, University of Oregon

Note: In addition to these sessions of invited papers, the American Economic Association, in cooperation with the Econometric Society, is also arranging for several sessions of contributed papers. The scheduled times and other details for these sessions will be given in the final program to be distributed at the meetings.

* Not to appear in the *Papers and Proceedings*

ANNOUNCEMENTS

Starting January 1, 1974, all manuscripts submitted to this *Review* have been sent out to referees without the author's name or affiliation. Original submissions should have a separate cover page for the title, author's name, and affiliation. The first page of the text should carry the title of the paper. For purposes of refereeing, authors are requested to avoid self-identification in the main text of their papers.

Resolutions for Consideration at the Annual Business Meeting

The Executive Committee at its meeting on March 8, 1974, voted to require that, to be considered at the annual business meeting, proposed resolutions must be submitted to the Secretary at least one month in advance in writing with signatures of the proposer and the second, both of whom must be members in good standing. The Secretary will reproduce the proposed resolutions and make copies available in advance of the meeting. The next business meeting will begin at 4:00 p.m. (please note the departure from the usual time) on December 29, 1974. The deadline for proposed resolutions is, accordingly, November 29. They should be sent to the Secretary, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

The Asia Foundation has awarded the American Economic Association a grant to assist graduate students and visiting professors from Asia currently at institutions west of the Mississippi to attend the annual meeting of the Association. The 1974 meeting will be held in San Francisco, California, December 28–30. Approximately eight travel grants up to $150 each will be awarded. Write the American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212 for an application form. To be considered for a travel grant, completed forms must be returned by October 31, 1974.

Annual Meeting Placement Service

The Placement Service at the 1974 annual meetings of the Allied Social Science Associations at the St. Francis Hotel in San Francisco will begin operation on December 27, the day before sessions begin. Applicants and employers will be able to attend more sessions with a day set aside entirely for labor market transactions. The Placement Service will continue during the other days of the annual meetings, but hopefully at a less frantic pace than in previous years. It will be open from 9:00 A.M. to 5:00 P.M., December 27; 8:30 A.M. to 5:00 P.M., December 28–29; and 8:30 A.M. to 12:00 noon, December 30.

For your convenience the National Registry for Economists is offering a preconvention registration service. Listing your application or position prior to the meetings will expedite service at the placement center. Applications and orders from individuals who are unable to attend the meetings will be available for review. Please report to the Placement Service immediately upon your arrival at the convention to activate your registration form. Forms may be obtained from the National Registry for Economists, Illinois State Employment Service, 40 West Adams Street, Chicago, Illinois 60603. They must be completed and returned to the National Registry prior to December 6, 1974.

Job Openings for Economists

In October the American Economic Association will beginp ublication of *Job Openings for Economists (JOE)*, a bimonthly listing of job openings. We strongly urge department chairmen, government employers, and members of the business community to list their vacancies in this publication. Subscriptions are available for the first year for $6 (domestic) and $11 (foreign). See the advertising section of this issue for an application form. Please note that only jobs are listed. Applicants should list themselves with the National Registry in Chicago. With the publication of *JOE*, we will discontinue the "Vacancies and Applications" section which normally appears in each regular issue of this *Review*. September 1974 is the last issue in which job openings will appear. December is the last issue for economists listing their availability for employment.

Innovations in Teaching Undergraduate Economics Courses

The American Economic Association Committee on Economic Education has reserved space at the December meetings to provide interested economics teachers with an opportunity to view, discuss, and have some hands-on experience with new technological innovations in teaching undergraduate courses. Further details will be included in the official program.

The purpose of this note is to alert the profession to the display and to invite any member who has an innovation to offer to submit details for consideration of possible inclusion. Unfortunately, the Association will be unable to provide any financial support beyond providing space and standard audio-visual equipment. Provision of any additional hardware (e.g. computer terminals and printed material) will be the responsibility of the contributor. Persons interested in submitting proposals should write to Professor Keith G. Lumsden, Graduate School of Business, Stanford University, Stanford, California 94305.

Visiting Economics Scholars Program

At its meeting on December 27, 1973, the Executive Committee of the American Economic Association approved a proposal for a Visiting Economics Scholars Program by Edward J. Powers of Northern Michigan University. The program will be modeled on the Visit-

ing Scientists Program formerly funded by the National Science Foundation and will begin with the academic year 1974–75. The goal is to provide visiting scholars to the more geographically isolated and less affluent colleges and universities across the country. For more information, contact C. Elton Hinshaw, Assistant Secretary-Treasurer, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

Income Tax Relief for Members in the United Kingdom

The Commissioners of Inland Revenue have approved the American Economic Association for the purposes of Section 192 Income and Corporation Taxes Act 1970 (formerly Section 16 Finance Act 1958). The conditions of the approval are as follows: "With effect from the year commencing 6 April 1973 a member who is assessable to Income Tax under Schedule E in respect of the emoluments of an office or employment is entitled to a deduction from those emoluments of the whole of the annual subscription which is due and payable by him to the body in the Income Tax year, provided that (a) the subscription is defrayed out of the emoluments of the office or employment, and (b) the activities of the society so far as they are directed to qualifying objects are relevant to the office or employment; that is to say, the performance of the duties of the office or employment, or the exercise of the profession concerned, is directly affected by the pursuance of the qualifying objects. A member of the body who is entitled to the relief should apply to his Tax Office as soon as possible giving details of his subscription and making a claim for the relief due to him."

The German Marshall Fund of the United States, an American private foundation, announces five to eight full-time *Marshall Fund Common Problems Fellowships* for 1975–76. They will be offered to outstanding U.S. scholars whose work is designed to contribute "to the better understanding and resolution of significant contemporary or emerging common problems of industrial societies," particularly in their comparative, international, social, political, and economic aspects. Write for information and application form (deadline Oct. 31, 1974) to the German Marshall Fund, 1717 Mass. Ave., N.W., Washington, D.C. 20036.

The Carnegie Endowment for International Peace announces a program of travel and maintenance support to pre- and postdoctoral scholars in the field of international organization. The subject area is "Changes in the international system that result in significant shifts in the locus of political authority and political allegiance away from the territorial state. At present, emphasis is being given to the following areas: 1) transnational relations, 2) elite networks, 3) impact studies." Address inquiries and applications to Robert C. Richter, Carnegie Endowment for International Peace, 345 East 46th Street, New York, New York 10017.

The Mathematics Research Center at the University of Wisconsin-Madison will hold a symposium on adaptive economics Oct. 21–23, 1974. The symposium will consist of three days of invited addresses and contributed papers dealing with the formal study, using mathematics and computer methods, of adaptive behavior, adaptive strategies, and evolutionary processes in economics and industrial settings. Members of the program committee are R. H. Day, chairman; T. Groves, J. Marschak, S. M. Robinson, and D. Rudd. Requests for the detailed program and information on registration and accommodations should be directed to Professor R. H. Day, Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706.

The Consortium for International Studies Education is sponsoring two one-week Learning Package Development Conferences on Transnational Problems, one of which will be held in Florida during the week of December 15. The second will be held on the West Coast during the week of January 5. The sites will be announced shortly. The purpose of the conferences is to generate a set of interdisciplinary learning packages which focus on specific transnational problems (e.g., world economic development and stability, management of world energy supplies and resources, control of violence, population), and have promise of widespread dissemination and usage in undergraduate curricula throughout the United States. The scholars invited to attend will be drawn from several disciplines on the basis of proposals submitted in a national competition. Awards will consist of a stipend, travel, and living expenses. Funds are also available to support the travel of instructors who are attending for training purposes and who are willing to field test the materials in their classrooms.

Individuals interested should submit a two-page proposal along with a vita *no later than November 1, 1974*, specifying the transnational problem focus of the package, the educational objectives, the learning "media" (data sets, A–V, simulation, etc.), evaluation mechanisms, and the anticipated length and cost of the package. Proposals and inquiries should be sent to the Project Director, J. Martin Rochester, Center for International Studies, University of Missouri-St. Louis, St. Louis, Missouri 63121.

Fellowships in Western Europe, 1975–76

Under a joint program of the Social Science Research Council and the American Council of Learned Societies, fellowships are offered for nine to eighteen months of dissertation research on contemporary Western European affairs. Applications are invited from students in all social sciences and humanities, but particular attention is given to disciplines in which relatively less attention has been devoted to Western Europe, such as anthropology, economics, social psychology, and sociology. Requests for further information and application forms should be sent to Social Science Research Council, Western European Program, 605 Third Avenue, New York, New York, 10016.

Harold Laski

I would like to hear from readers who knew Professor Harold Laski, the British-born lecturer in political science who spent a substantial part of his professional life in the United States. For nearly one-third of his teaching career, he was associated with American universities and colleges and made many friendships which lasted over the years.

I am engaged on a study of Laski's life and work (he died in 1950) and am seeking personal reminiscences, recollections of his personality, and assessments of his influence and work. Please write to Granville Eastwood, 16, The Vineries, Enfield, Middlesex EN1 3DQ, England.

The Southern Regional Science Association will hold its 1975 meetings in Atlanta, Georgia, Apr. 3–4. Scholars interested in presenting papers on any subject involving analysis of the spatial dimensions of human activity are invited to submit one-page abstracts by Oct. 1, 1974 to the program chairman, Professor William H. Miernyk, Regional Research Institute, 4 Main Library, West Virginia University, Morgantown, West Virginia 26506.

A conference on "Current Research in the Economics of Education" will be held at the Educational Testing Service in Princeton, New Jersey on Friday and Saturday, Oct. 18–19, 1974. The purpose of this conference is to improve communication among economists working in the field through a set of informal workshop sessions. Anyone who would like to participate should contact Dr. Dean Jamison at E.T.S. or Professor Paul Wachtel, New York University Graduate School of Business.

Deaths

John A. Doukas, associate professor of economics, Miami University, Jan. 10, 1974.

Victoria C. Lapham, research economist, Urban Institute, Washington, Jan. 20, 1974.

Edward E. Lewis, Howard University, retired, Silver Springs, Maryland, Mar. 10, 1974.

Edwin G. Nourse, first chairman of the President's Council of Economic Advisers, Bethesda, Maryland, Apr. 7, 1974.

Paul N. Taylor, professor, department of economics, University of Connecticut, Apr. 7, 1974.

Winfield W. Riefler, former assistant to the chairman of the Federal Reserve Board, retired, Sarasota, Florida, Apr. 5, 1974.

Retirements

Rolfe A. Haatvedt, professor of economics and business administration, Luther College, Aug. 31, 1973.

William T. Hicks, chief, Economics Branch, Mississippi River Commission, U.S. Army Corps of Engineers, Jan. 1974.

Visiting Foreign Scholars

Christopher I. Higgins, Australian Treasury: visiting associate professor of economics, University of British Columbia, fall 1974.

Wan-Soon Kim, Korea University: visiting professor of economics, American Graduate School of International Management, spring 1974.

David T. Lapkin, University of Haifa, Israel: visiting professor of economics, New Mexico State University, 1974–75.

David Levhari, Hebrew University: visiting professor in economics, University of Illinois, June–Dec. 1974.

S. B. Ngcobo, University of Botswana, Lesotho and Swaziland: visiting lecturer, Economic Development Institute of the World Bank, Jan.–June 1975.

Richard Perlman, University of Wisconsin, Milwaukee: visiting professor, department of economics and commerce, Simon Fraser University, May 1, 1974.

Hans Ruthenberg, Lehrstuhl fur Okonomik der Làndwirtschaft, Germany: visiting lecturer, Economic Development Institution of the World Bank, Sept.–Dec. 1974.

Brinley Thomas, University College, Wales: visiting professor in economics, University of Illinois, spring 1975.

Promotions

Alan Abouchar: professor, department of political economy, University of Toronto, July 1974.

Mary Ann Bailey: assistant professor of economics, Yale University, Jan. 1974.

David P. Baron: professor, department of managerial economics and decision sciences, Northwestern University, Sept. 1974.

Nancy S. Barrett: professor of economics, American University.

Thomas Birnberg: associate professor of economics, Yale University, July 1974.

Jean-Marie Blin: associate professor, department of managerial economics and decision sciences, Northwestern University, Sept. 1974.

Joseph T. Bomelles: professor, department of economics, John Carroll University, Sept. 1974.

Donald J. Brown: associate professor of economics, Yale University, July 1974.

Chu-yuan Cheng: professor, department of economics, Ball State University, Sept. 1, 1974.

Benjamin G. Davis: senior associate, Education Division, RMC Research Corporation, Feb. 1, 1974.

Ross D. Eckert: associate professor of economics, University of Southern California, July 1, 1974.

Teviah L. Estrin: assistant professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1973.

Edgar L. Feige: professor, department of economics, University of Wisconsin-Madison.

John D. Ferguson: assistant professor of economics, Miami University, Mar. 25, 1974.

William F. Ford: executive director, research and planning; chief economist, American Bankers Assoc., Jan. 14, 1974.

Edward T. Howe: assistant professor, Siena College, Sept. 1974.

Boris M. Ivezic: professor of economics, Waynesburg College, Feb. 1974.

Michael J. Lavelle: associate professor, department of economics, John Carroll University, Sept. 1974.

Patrick C. Mann: professor, department of economics, West Virginia University, Aug. 1974.

Samuel Morley: associate professor, department of economics, University of Wisconsin-Madison.

Neil A. Palomba: professor, department of economics, West Virginia University, Aug. 1974.

David B. Pariser: associate professor, department of economics, University of North Dakota, Aug. 1974.

John M. Quigley: associate professor of economics, Yale University, July 1974.

Donald J. Roberts: associate professor, department of managerial economics and decision sciences, Northwestern University, Sept. 1974.

Robert L. Robertson: professor, department of economics, Mount Holyoke College, July 1, 1974.

John Rapoport: associate professor, department of economics, Mount Holyoke College, July 1, 1974.

Larry B. Sawers: associate professor of economics, American University.

Stephen J. Schmutte: associate professor, department of economics, Wabash College, Sept. 1974.

Ross M. Starr: associate professor of economics, Yale University, July 1974.

Michael K. Taussig: professor of economics, Rutgers-The State University, July 1, 1974.

Walter J. Wadycki: associate professor of quantitative methods, University of Illinois at Chicago Circle.

Gregory J. Yi: associate professor, department of economics, West Virginia University, Aug. 1974.

Administrative Appointments

Nancy S. Barrett: chairman, department of economics, American University.

Warren G. Berg: dean of academic affairs, Luther College, Sept. 1, 1973.

Robert J. Lampman: chairman, department of economics, University of Wisconsin-Madison.

Michael J. Lavelle: chairman, department of economics, John Carroll University, Sept. 1973.

Stanton C. Lindquist: chairman, department of economics and business administration, Luther College, Sept. 1, 1973.

M. Gene Newport: dean, School of Business, University of Alabama, Birmingham.

James W. Nordyke: acting department head, department of economics, New Mexico State University, 1974–75.

Neil A. Palomba: director, Labor Relations Institute, West Virginia University, July 1974.

Harl E. Ryder: chairman, department of economics, Brown University, July 1, 1974.

Joseph J. Seneca: chairman, department of economics, Rutgers-The State University, July 1973.

Victor E. Smith: chairman, department of economics, Michigan State University, Sept. 1, 1974.

Nancy Z. Spillman: director, Center for Economic Education, Los Angeles Trade-Technical College, Mar. 15, 1974.

F. M. Scherer, International Institute of Management, Berlin: director, Bureau of Economics, Federal Trade Commission, July 1974.

James Tobin: chairman, department of economics, Yale University, July 1974.

Jerry D. Young: vice president for finance, University of Alabama, Birmingham.

Raymond E. Zelder: chairman, department of economics, Western Michigan University, July 1, 1974.

Appointments

Masanao Aoki: professor of economics and engineering, University of Illinois, 1974–75.

Lloyd C. Atkinson, University of Maryland: associate professor of economics, Miami University, Sept. 1974.

Werner Baer: professor of economics, University of Illinois, fall 1974.

Robert Bloom: lecturer, department of economics, Rutgers-The State University, Jan. 1974.

Eugene E. Bond: associate professor, world business department, American Graduate School of International Management, summer 1973.

Allan R. Cahoon, Syracuse University: assistant professor, faculty of business, University of Calgary, July 1, 1974.

Mathew Canzoneri: assistant professor of economics, University of Illinois, 1974–75.

Charles Cicchetti: associate professor, department of economics, University of Wisconsin-Madison.

Sheldon Danziger: research associate, Institute for Research on Poverty, University of Wisconsin-Madison.

Michael R. Darby: research fellowship, National Bureau of Economic Research, 1974–75.

Murray L. Davis, Wilfrid Laurier University: associate professor, faculty of business, University of Calgary, July 1, 1974.

Lawrence E. Demilner: instructor in economics, University of Southern California, Sept. 1, 1974.

Jonathan G. Dickinson: research associate, Institute for Research on Poverty, University of Wisconsin-Madison.

Sergei P. Dobrovolsky, Rensselaer Polytechnic Institute: visiting professor, faculty of management, McGill University, 1974–75.

James J. Doyle: assistant professor, department of economics, John Carroll University, Sept. 1974.

Ray C. Fair, Princeton University: associate professor of economics, Yale University, July 1974.

Lyle E. Fogel: professor of economics, Wabash College, Sept. 1974.

Gary Fromm: senior research staff, National Bureau of Economic Research, Jan. 1, 1974.

Adam Gifford, Jr., California State University at Northridge: assistant professor of economics, Southern Methodist University, 1974–75.

Frank Gollop: assistant professor, department of economics, University of Wisconsin-Madison.

David K. Guilkey, University of North Carolina:

assistant professor of economics, Southern Methodist University, 1974–75.

William R. Hart, Washington University: assistant professor of economics, Miami University, Sept. 1974.

Kichiro Hayashi: associate professor, American Graduate School of International Management, fall 1974.

James S. Henderson, Riddell, Stead & Co.: assistant professor, faculty of business, University of Calgary, Sept. 1974–Apr. 1975.

J. Vernon Henderson, Queen's University: assistant professor, department of economics, Brown University, July 1, 1974.

Bryce Hool: assistant professor, department of economics, University of Wisconsin-Madison.

Ming-jeng Hwang: assistant professor, department of economics, and research associate, Regional Research Institute, West Virginia University, Aug. 1974.

Yannis M. Ionnides, University of California, Riverside: assistant professor of economics, Brown University, July 1, 1974.

James Jonish, University of Hawaii: associate professor of economics, Texas Tech University, Sept. 1973.

Robert E. Kalaba: professor of engineering and economics, University of Southern California, Sept. 1, 1974.

Barry P. Keating, University of Notre Dame: assistant professor, department of economics, Virginia Polytechnic Institute, Sept. 1974.

Jonathan R. Kesselman: visiting assistant professor in economics and research associate, Institute for Research on Poverty, University of Wisconsin-Madison, 1974–75.

Richard Kihlstrom: associate professor of economics, University of Illinois, 1974–75.

Taeho Kim: assistant professor, world business department, American Graduate School of International Management, spring 1973.

David Kotz: assistant professor of economics, American University.

Elisabeth M. Landes: research associate, National Bureau of Economic Research, Jan. 21, 1974.

George F. Lane, Sir George Williams University: associate professor, faculty of business, University of Calgary, July 1, 1974.

Wu-Lang Lee: research associate, National Bureau of Economic Research, Jan. 1, 1974.

Duanne Leigh, Washington State University: visiting associate professor, Institute for Research on Poverty, University of Wisconsin-Madison.

Richard E. Levin: lecturer in economics, Yale University, July 1974.

Lucinda M. Lewis: assistant professor, department of economics, University of Wisconsin-Madison.

Vincent L. Y. Lin, International Bank for Reconstruction and Development: assistant professor of economics, St. Mary's University, Aug. 1974.

John R. Lyons: assistant professor of economics, American University.

Daniel L. McDonald, University of Alberta: professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1974.

Harvey J. McMains: senior research associate, National Bureau of Economic Research, Feb. 19, 1974.

Michael S. McPherson, University of Chicago: assistant professor of economics, Williams College, July 1, 1974.

G. S. Maddala: research associate, National Bureau of Economic Research, Jan. 19, 1974.

Nestor Marquez-Diaz, Texas A & I University, Laredo: director of economic research, Economic Development Administration, Commonwealth of Puerto Rico, Continental Operations Branch.

James W. Meehan, Jr.: assistant professor of economics, Colby College.

John A. Menefee, Duke University: assistant professor of economics, California State College, Bakersfield, Sept. 1974.

Jerome W. Milliman, University of Southern California: professor of economics, University of Florida, Sept. 1, 1974.

Leonard Mirman: professor of economics, University of Illinois, 1974–75.

Malcolm C. Munro, University of Minnesota: assistant professor, faculty of business, University of Calgary, Sept. 1, 1974.

Frank J. Navratil: assistant professor, department of economics, John Carroll University, Sept. 1973.

Carl R. Neu, Harvard University: associate economist, economics department, The Rand Corporation, Sept. 1974.

Ubadigho Okonkwo: assistant professor, departments of economics, Afro-American studies, Institute for Research on Poverty, University of Wisconsin-Madison.

Massahiro Okuno: assistant professor of economics, University of Illinois, 1974–75.

Donald C. Oliver: research associate, National Bureau of Economic Research, Jan. 23, 1974.

Sharon M. Oster, Harvard University: lecturer in economics, Yale University, July 1974.

Walter P. Page: assistant professor, department of economics, West Virginia University, Aug. 1974.

Mark V. Pauly: research fellowship, National Bureau of Economic Research, 1974–75.

William Peterson: professor, American business, American Graduate School of International Management, spring 1974.

William Poole, Board of Governors, Federal Reserve System: professor of economics, Brown University, July 1, 1974.

Andrew Postlewaite: assistant professor of economics, University of Illinois, 1974–75.

Kathleen Powers: assistant professor, world business department, American Graduate School of International Management, spring 1974.

Rama V. Ramachandran, Brown University: assistant professor of economics, Southern Methodist University, 1974–75.

Nallapu N. Reddy: associate professor of economics, University of Michigan-Flint, Sept. 1974.

George F. Rhodes, Texas A&M University: assistant professor, Ohio State University.

William H. Rogers, Stanford University: associate

statistician, economics department, The Rand Corporation, Jan. 1975.

Efraim Sadka: assistant professor, department of economics, University of Wisconsin-Madison.

Steven H. Sandell, Central Bank of Ireland: assistant professor, Ohio State University.

Steven Schrader: instructor, department of economics and business administration, Luther College, Sept. 1, 1974.

Robert J. Shiller: research fellowship, National Bureau of Economic Research, 1974–75.

Brock K. Short, University of Windsor: economist, International Monetary Fund, May 15, 1974.

James P. Smith, City University of New York and National Bureau of Economic Research: associate economist, economics department, The Rand Corporation, Sept. 1974.

Rodney T. Smith, University of Chicago: associate economist, economics department, The Rand Corporation, Sept. 1974.

Vernon L. Smith, California Institute of Technology: professor of economics, University of Southern California, Sept. 1, 1974.

Richard H. Steckel, University of Chicago: assistant professor, Ohio State University.

Marvin J. Sternberg, State University of New York, Albany: research associate, department of economics, University of California, Berkeley, 1974–75.

Daniel B. Suits, University of California, Santa Cruz: professor, department of economics, Michigan State University, Sept. 1, 1974.

Amemiya Takeshi: research associate, National Bureau of Economic Research, Jan. 19, 1974.

William A. Tilleman, University of Wisconsin: associate professor, faculty of business, University of Calgary, July 1, 1974.

Andris Trapans: assistant professor, international studies department, American Graduate School of International Management, spring 1973.

Roger C. Vergin: professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1974.

William P. Wadbrook, University of Maryland: professor of economics, Industrial College of the Armed Forces, June 1, 1974.

Robin J. Walther: instructor in economics, University of Southern California, Sept. 1, 1974.

Eugene H. Warren, Jr., University of Tennessee: assistant professor, faculty of business, University of Calgary, Sept. 1, 1974.

Charles A. Wilson, University of Rochester: lecturer in economics, Yale University, July 1974.

Jon D. Wisman: assistant professor of economics, American University.

John Yinger: research associate, Institute for Research on Poverty, University of Wisconsin.

Leaves for Special Appointments

Ross D. Eckert, University of Southern California: research fellow, Hoover Institution on War, Revolution, and Peace, Stanford University, Sept. 1, 1974–75.

Robert W. Gillespie, University of Illinois: visiting fellow, National Institute of Law Enforcement and Criminal Justice, Washington, D.C., Aug. 1974–Jan. 1975.

Randall A. Hoffman, Jr., Iowa State University: leader of farm management and production economics component, Iowa Universities Peru program in Lima, Jan. 1974–76.

Jane H. Leuthold, University of Illinois: Ford Foundation faculty fellowship on the role of women in society, 1974–75.

Jose Mulleady, Iowa State University: farm management economist, Iowa-Peru project in Lima, Feb. 1974–76.

Arturo C. Porzecanski, University of Pittsburgh: visiting scholar, Brookings Institution, Aug.–Sept. 1974.

Walter J. Primeaux, Jr., University of Mississippi: visiting professor of business administration, University of Illinois, 1974–75.

Donald J. Roberts, Northwestern University: research associate, Center for Operations Research and Econometrics, Belgium, Sept. 1974–June 1975.

Gian S. Sahota, Vanderbilt University: research project, Institute of Economic Growth, Delhi, summer–fall 1974.

Ryuzo Sato, Brown University: Institute of Economic Research, Hitotsubashu University, Sept. 1974–Jan. 1976.

Mark B. Schupack, Brown University: visiting scholar, Massachusetts Institute of Technology, Aug. 1974–July 1975.

Geoffrey S. Shepherd, Iowa State University: coordinator, Inter-American Development Bank, Paraguay, Jan. 1974–June 1975.

David G. Tarr, Ohio State University: economist, Bureau of Industry Analysis, Federal Trade Commission, Washington.

Thomas H. Tietenberg, Williams College: Brookings Economic Policy Fellow to the Federal Energy Office, July 1974–Sept. 1975.

Resignations

Donald W. Bryan, University of Calgary, June 30, 1974.

Robert E. Evenson, Yale University: Agricultural Development Council, July 1974.

Claudia D. Goldin, University of Wisconsin-Madison: Princeton University, Sept. 1974.

Theodore Groves, University of Wisconsin-Madison: Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Tatsuo Hatta, Ohio State University: Saitama University, Japan, Sept. 1974.

George A. Hay, Yale University: U. S. Department of Justice, July 1974.

Edward H. Murphy, University of Calgary, June 30, 1974.

Wallace E. Ogg, Iowa State University, Dec. 31, 1973.

Rudolph K. Schnabel, Iowa State University: University of Arizona, Nov. 1973.

Joseph E. Stiglitz, Yale University: Stanford University, July 1974.

William P. Travis, University of California, San Diego: University of Indiana, Jan. 1, 1974.

Edwin M. Truman, Yale University: Board of Governors, Federal Reserve Stystem, July 1974.

Henry C. Wallich, Yale University: Board of Governors, Federal Reserve Board, Mar. 1974.

Richard Weisskoff, Yale University: Iowa State University, July 1974.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, John W.), his present place of employment or enrollment: his new title (if any), his next place of employment (if known or if changed), and the date at which the change will occur.

C. Type each item on a separate 3×5 card, and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, November 1; *June*, February 1; *September*, May 1; *December*, August 1.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office. All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

NATIONAL REGISTRY FOR ECONOMISTS

The National Registry for Economists was established in January, 1966, to provide a centralized nationwide clearinghouse for economists on a year-round basis. It is located in the Chicago Professional Placement Office of the Illinois State Employment Service and is staffed by experienced placement personnel, operating under the guidance and direction of Regional and National Bureau of Employment Security Professional Placement officials, and in cooperation with the American Economic Association. It is a free service. The National Registry for Economists maintains completely separate listings from those of the American Economic Association, and the National Registry does *not* have the listings as shown in the *American Economic Review*. There are no registration, referral, or placement fees. Application and order forms used in the Registry are available upon request from the: National Registry for Economists, Professional Placement Center, 40 West Adams Street, Chicago, Illinois 60603.

AMERICAN ECONOMIC ASSOCIATION
VACANCIES AND APPLICATIONS
To Be Phased Out

In October 1974 the American Economic Association will begin publication of *Job Openings for Economists* (*JOE*), a bimonthly listing of job vacancies. Subscriptions to *JOE* will be available to members of the Association, departments of economics, and other organizations at an introductory rate of $6 (USA, Canada, Mexico, first class) or $11 (foreign airmail) for the first year. For more information, write *JOE*, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212. Please note that only *jobs* will be listed.

The Vacancies and Applications section of the *American Economic Review* will be phased out as the publication of *JOE* begins. Vacancies will be published for the last time in the September 1974 issue. Applications will be published for the last time in the December 1974 issue. See a previous issue for details on preparing and answering listings.

Vacancies

Research Assistants: The Institute for Research on Human Resources at The Pennsylvania State University has a recurring need for research assistants to fill noncontinuing, nontenured positions.

Economist: The Department of Economics at Grand Valley State Colleges is recruiting an Instructor for the fall term (September 26-December 13) of the 1974-75 academic year. Applicants should have completed all requirements for the Ph.D. except dissertation. Teaching experience of some type in principles of microeconomics is necessary. Duties consist of teaching one section of principles of microeconomics and one section of either public finance or money and banking. Salary is $4,000. GVSC is an equal opportunity employer. Send letters of application and résumés to John W. Reifel, Chairman, Department of Economics, Grand Valley State Colleges, Allendale, Michigan 49401.

Distinguished Economists: As part of a major building program the Department of Economics at the University of Southern California is seeking several distinguished economists with outstanding national reputations. A substantial list of publica-

tions in major journals is required. For further details write (enclosing detailed résumé) to John H. Niedercorn, Chairman, Department of Economics, University of Southern California, Los Angeles, California 90007.

Economists: Senior Professor, and lower level appointments available. Applicant would join a 45 member, interdisciplinary school of social sciences with faculty in economics, anthropology, geography, political science, psychology, and sociology. Applications from highly qualified candidates whose ability and interests are broad and cut across traditional boundaries are welcome. Send vita, list of three references and short description of research and teaching plans to Dean L. A. Forman, School of Social Sciences, University of California, Irvine, California 92664. Applications from all qualified candidates are welcome; minorities and women are encouraged to apply.

Economists: New Midwest research institute seeks Ph.D. radical macroeconomists for semi-scholarly research on war-peace reconversion, economics of health care, etc. Prefer unselfish, socially conscious, non-careerist, Phi Beta Kappa, previous Movement work, publications, ability to get grants, etc. Salary $10K-$20K. P378

Economics, particularly with Science/Engineering Complement: A medium-sized Washington-based applied research firm specializing in public policy research. Requirements are interest in problem solving research, high aptitude, training in principles of economics, and subject matter knowledge in energy, environment, or transportation. Salary open. Send résumés to Jack Faucett Associates, Inc., 5454 Wisconsin Avenue, Suite 1150, Chevy Chase, Maryland 20015. Equal opportunity employer.

Economists: Haifa University's new and expanding economics department seeks economists in a wide range of specializations for its current undergraduate and planned graduate programs. Contact Dr. Dan Eldor, Chairman, Department of Economics, Haifa University, Israel.

Regional-Urban Economics and Statistics: The Department of Economics will have an opening January 1975 for a faculty member in the fields of regional-urban economics and statistics. The Department seeks an economist with strong technical competence and dedication to teaching and research. An appointment at the Assistant Professor level (salary approximately $12,000-$13,500) but appointments above this level will be considered. Applicant must have Ph.D. For further information and details, write Larry J. Larsen, Chairman, Department of Economics, University of Nevada, Reno, Reno, Nevada 89507. Please enclose a detailed résumé. The University of Nevada is an affirmative action—equal opportunity employer.

Economists/Geographers: Applications are invited for a Fellowship or Fellowships tenable at any level up to post-doctoral, for research on a topic related to off-shore petroleum/gas exploitation. The interest of the Institute is focused on the social and economic life of Newfoundland and its region. Applications to: Secretary, Institute of Social & Economic Research, Memorial University of Newfoundland, St. John's, Newfoundland, A1C 5S7.

Extension Marketing Economist: A full-time position is open for a faculty member of the University of Georgia College of Agriculture State Extension Staff. Requires Ph.D. in agricultural economics, academic training and/or experience in grain marketing, livestock marketing, and/or transportation. Ability to conduct educational programs in marketing economics to include feasibility studies, market analysis, outlook, workshops and seminars with grain, livestock and/or transportation clientele and county Extension agents. Must be U.S. citizen. Contact: Dr. James B. Harris, Head, Extension Personnel Development Department, University of Georgia, 216 Extension Building, Athens, Georgia 30601, or call 404/542-2713.

Economics: Applications are invited for a position at the assistant professor level, January 1975. Public finance, monetary and development are the desired teaching areas. Applicants must have the Ph.D. degree. Write: Chairman, Department of Economics, University of Alaska, Anchorage, Anchorage, Alaska 99504.

Professors of Economics: Applications are invited for three posts, including two full professorships, one carrying the Chairmanship of the Department. Interests in any field of economics other than economic history, but including econometrics, labour economics and industrial relations. Salary commensurate with qualifications and experience. Send curriculum vitae and names of three references to Registrar, University of Otago, Dunedin, New Zealand, by 31st January, 1975.

Economics: Applications are invited for a one-year appointment. Teaching areas: principles of economics plus two of the following: medical care, statistics, price theory, Soviet Union, poverty. Ph.D. (or near) preferred. Rank and salary commensurate with qualifications and experience. Applications should be sent to Dr. Robert Bowers, Head, Department of Economics, Western Michigan University, Kalamazoo, Michigan 49001. An equal opportunity affirmative action employer.

Economics: Assistant professor, Ph.D., Department of Economics, University of the Pacific, Stockton, California. Primarily to teach courses in comparative economic systems and economic development. However, because of the size of the department, teaching competencies in some other areas of economics are not only considered desira-

ble but necessary. Because of a unique freshman year program in the college, some experience in interdisciplinary pursuits or a willingness to explore interdisciplinary themes is also necessary. Occasional participation in the freshman year program is expected. Address résumés to: Professor John P. Carew, Chairman, Department of Economics, University of the Pacific, Stockton, California 95211.

Economists: Senior vacancies for economists at OECD Paris, in the Department of Economics and Statistics. *Deputy Director of the General Economics Branch:* The Deputy Director shares with the Director the task of stimulating and supervising analysis by the various Divisions which deal with trade and payments relationships, domestic monetary developments and international capital flows, longer-term growth and resource allocation problems, inflation and related policy issues, and any other general economic questions that may be of concern to the OECD. The successful candidate will be capable of exploiting work based on modern analytical techniques, will have had experience of, and a feeling for, the problems of economic policy, and will have the ability to present issues, in writing and orally, in a form which official policy-makers will find useful. He will be able to co-ordinate the work of some 25 professional economists of different nationalities. Basic salary range: 118,000 to 149,000 French francs. *Head of Monetary Division:* The Division is responsible for analysis, for the Economic Policy Committee and its Working Party No. 3, of domestic and external monetary policies of member countries and the preparation of papers for the Committee for Monetary and Foreign Exchange Matters on developments in foreign exchange markets and their inter-relation with domestic monetary policies and controls on international capital movements. *Head of Economic Prospects Division:* The Division is responsible for work undertaken on behalf of the Economic Policy Committee for the twice yearly review of general economic prospects. The work involves identification and analysis of shorter-term developments in Member countries which have possible inter-country repercussions, and the presentation of these to the Economic Policy Committee. The Division is also responsible for the preparation and publication of the twice yearly Economic Outlook, the Occasional Studies Series and other published material, and carries out research on new methods of forecasting and analysis. Basic salary range for Head of Division: 100,000 to 137,000 French francs. Expatriation allowances for employees not normally resident in France plus other allowances add 16 percent to 26 percent to basic salary. Emoluments are normally tax free. Applications from nationals of OECD Member countries, accompanied by detailed curricula vitae in English or French (the official languages of OECD) should be addressed to Personnel Division, Organisation for Economic Co-operation and Development, 2 rue André Pascal, Paris 75016.

Faculty of Economics, The University of West Florida: Applications are invited for one teaching position with competencies required in macroeconomics, development, international and urban economics. Includes teaching in our graduate program. Assistant professor. Ph.D. required. Salary based on qualifications and experience.. Write to Dr. Pedro C. M. Teichert, Chairman, Faculty of Economics, The University of West Florida, Pensacola, Florida 32504. The University of West Florida is an equal employment and educational opportunity institution.

Assistant Professor of Economics: Regular, full-time economics faculty position in the Social Science Division, University of Colorado at Denver, for September 1975. The applicant should have a Ph.D. or expect to have the degree conferred by September 1974, and should have some teaching experience. Applications are particularly encouraged from women and persons from racial and ethnic minorities. Desired research and teaching fields are transportation, human resources, natural resources and environmental economics, industry organization, medical economics, and econometrics. Persons with other specialties should not apply. The position offers opportunities to participate in interdisciplinary teaching and research programs. Preference will be given to persons whose research focuses on urban problems. Teaching includes undergraduate and masters level courses. Rank and salary for a Ph.D. are assistant professor and $13,000-$15,000 respectively; persons still completing their dissertations will be considered for an instructorship at a lower salary only. The University is an equal opportunity employer. Write Professor Suzanne W. Helburn, Economics Chairperson, University of Colorado at Denver, 1100 14th Street, Denver, Colorado 80202. Include résumé and references.

Economist IV: The Alberta Hospital Services Commission requires a senior economist to conduct complex and sophisticated analyses and projections of the Commission's component of the Health Care System and to provide the Commission with information on economic aspects of various programmes. To act in a consulting capacity and work in close cooperation with research officials from government, hospitals and divisions of the Commission. A masters degree in economics is required with several years experience in the health and hospital field in work related to economics, statistics and research programmes. Please send application to: R. D. LaRiviere, Director, Administrative Services and Personnel, Alberta Hospital Services Commission, P.O. Box 2222, Edmonton, Alberta, Canada.

Regional-Urban Economics and Statistics: The Department of Economics will have an opening August 1975 for a faculty member in the fields of regional-urban economics and statistics. The Department seeks an economist with strong technical competence and dedication to teaching and

research. An appointment at the assistant professor level (salary approximately $12,000-$13,500) but appointments above this level will be considered. Applicant must have Ph.D. For further information and details, write Larry J. Larsen, Chairman, Department of Economics, University of Nevada, Reno, Reno, Nevada 89507. Please enclose a detailed résumé. The University of Nevada is an affirmative action-equal opportunity employer.

Pharmacy Administration: A candidate is sought to teach pharmacy administration at the undergraduate and graduate levels and to conduct research alone and in a team. The position requires a research background in one of the social sciences and the ability to apply these skills to problems dealing with health care delivery with particular emphasis on pharmacy and the drug use process. Interested parties should send a c.v. to: Dr. Albert I. Wertheimer, College of Pharmacy, University of Minnesota, Minneapolis, MN 55455.

Energy Economics: The Center for Building Technology of the National Bureau of Standards invites applications for a position in the Building Economics Section. People are being sought who wish to apply economic theory to the solution of a broad range of energy problems faced by the construction industry. Responsibilities include proposal writing, contact with project sponsors, project leadership, and the writing of research reports to be published. Supervision is over all project personnel. A Ph.D. in economics with a strong background in microeconomics is required. Research and other experience will determine salary range between a grade of GS-12 at $17,497 per annum and a GS-13 at $20,677. Send Personnel Qualification Statement (Federal Standard Form 171) to Dr. Harold E. Marshall, Building 226, A-355, National Bureau of Standards, Washington, D.C. 20234. If you have any questions, please call (301) 921-3701.

Economics: The Department of Economics of Lowell Technological Institute anticipates several positions during the 1974-75 academic year at the instructor and assistant professor levels for economists holding their Ph.D. or in the final stages of their dissertations. Although several areas of specialization are open, the candidate should have a strong background in statistics and econometrics. Salaries are competitive and based on qualifications and experience. Lowell Technological Institute is a state supported institution located twenty-five miles northwest of Boston, Mass. As of 1975, it will function as the University of Lowell as a result of its merger with nearby Lowell State College. The Department of Economics is housed in the College of Management Science and attracts students from business, engineering, humanities, and science programs. Lowell Technological Institute is an equal opportunity employer and actively seeks women and minority candidates. Please send applications to: Dr. Thomas G. Macbeth, Hiring Committee Chairman, Lowell Technological Institute, Lowell, Mass. 01854.

Business Administration and Economics: The School of Business at the University of North Alabama anticipates several new openings in the Fall of 1975. Applications are invited for graduate and undergraduate teaching positions in accounting, marketing, management, finance, economics and quantitative management science. Teaching assignments are flexible. Qualifications required are Ph.D. or D.B.A. Employment would be at the rank of assistant or associate professor depending on qualifications. Starting salaries range from $15,000 up, depending on education and experience. The University of North Alabama is an equal opportunity employer and is actively seeking women and minority candidates. Please write to Dean Lawrence H. Conwill, School of Business, University of North Alabama, Florence, AL 35630.

Economics: Small midwest liberal arts college of the University of Minnesota, seeks DBA, Ph.D. or candidate to teach business related courses in an economics discipline. We are seeking a person committed to teaching excellence to provide course offerings from some of the following areas: accounting, finance, investment and managerial economics. Appointment begins September 16, 1975. Campus located in Morris, Minnesota (population 5,366), 150 miles from Minneapolis metropolitan area. Contact: Sun M. Kahng, Division of Social Sciences, University of Minnesota, Morris 56267.

Economics: Position open for an assistant professor with Ph.D. and some teaching experience for September, 1975. Background in monetary theory. Applicants with secondary fields in such areas as environmental, transportation, health economics or other current applied fields will be given special consideration. Primary emphasis on superior undergraduate teaching. 12 hours semester teaching load. Salary range $12,883-$15,565, dependent upon experience and qualifications. Applications close February 1, 1975. An on-campus interview at the applicant's expense is required. Write: Dr. Armand J. Zottola, Chairman, Department of Economics, Central Connecticut State College, New Britain, Connecticut 06050. Central Connecticut State College is an affirmative action equal opportunity employer.

Assistant Professor: The University of Texas at El Paso is seeking applicants for a full-time position beginning in January or September, 1975. The position to be filled will require teaching responsibility for teaching finance and micro or macroeconomic theory in a department of economics and finance. Applicants should have completed the terminal degree or have it clearly in sight by September, 1975. The University of Texas at El Paso is an equal opportunity employer. Applications from women and persons of minority or ethnic background are invited. Write Paul D. Zook, Chairman, Department of Economics and Finance, The University of Texas at El Paso, 79968.

Economist: The Office of Research and Development of the U. S. Department of the Interior seeks an economist to assist in the formulation and analysis of strategies for research and development in energy and minerals, with particular reference to fossil fuel research of the Bureau of Mines and Office of Coal Research. Although a demonstrated capacity for independent thought and critical review is more important than specific experience, background in resource economics and familiarity with energy markets will prove helpful. This is a career position in the Civil Service. The Office is also interested in a one-year appointment for an economist on sabbatical. Write to Mr. Nicolai Timenes, Jr., Assistant Director—Environment and Economic Analysis, Office of Research and Development, Room 4458, Department of the Interior, Washington, D.C. 20240.

Economist: The Department of Economics wishes to recruit an economist specializing in management science plus urban economics. Assistant professor or above. Salary open but not less than $13,930. Ph.D. required. Undergraduate and graduate teaching. 9 hour teaching load. Applications from women and members of minority groups are welcomed. Write to: Morris Silver, Chairman, Department of Economics, City College of New York, Convent Avenue at 138th Street, New York 10031.

Economist (Regional Economic Planner): A recently established regional planning commission seeks a economist to pursue an economic analysis and development program for the region. The successful applicant will be responsible for the development of a 14-month study of the regional economy, economic indicators, public investment strategies and the coordination of economic studies with other ongoing regional planning programs including a coastal zone management program and proposed Economic Development District. Qualifications should include a master's degree in regional planning, regional economics or a related field as well as a familiarity with regional planning and two years of demonstrated practical experience in the area of economic analysis. A working knowledge of computer application is necessary. Beginning salary range, $11,500 to $13,000 with liberal fringe benefits. Reply to Ralph M. Bergman, Executive Director, Bay-Lake Regional Planning Commission, 100 N. Jefferson Street, Green Bay, Wisconsin 54301. (414)432-7286.

Public Finance Economist: The Oregon Department of Revenue is taking applications for economists with extensive backgrounds in state and local public finance. The position requires a minimum of four years' experience in economic research. Applicants should have at least a master's degree with graduate level work in public finance. The position involves collection and analysis of data, planning basic research programs for use in the formulation of new legislation, evaluation of existing legislation, and general public finance research for legislative committees. Knowledge of advanced statistical research techniques and computer programming is desirable. The salary range is $13,968 to $17,844 annually. The State of Oregon is an affirmative action/equal opportunity employer and is actively seeking women and members of minority groups. Send résumés to Personnel Division, State of Oregon, Salem, Oregon 97310.

Micro and Macroeconomics: Positions available beginning in September, 1975 for recent Ph.D. graduates to teach in both undergraduate and graduate programs. Emphasis shall be placed upon quality instruction and continued professional growth, initially exemplified by some publication record in the area of specialization. Nationally competitive salaries. Send résumé to: Dr. Kurt F. Flexner, Chairman, Department of Economics, Memphis State University, Memphis, Tennessee 38152.

Researchers: The Batelle-Columbus Laboratories have openings for researchers in several areas, including market, new product, or corporation planning, international economic trends, East-West trade, role of the multi-national corporation, and transportation. Education requirements range from graduate work to Ph.D. Several years experience in related job required for most positions. Various locations in U.S. and some international travel. Contact: Philip L. Morris, Senior Personnel Advisor, Batelle-Columbus Laboratories, 505 King Avenue, Columbus, Ohio 43201.

Senior Lecturers in Managerial Economics and Health Economics: Applications are invited for appointment to the above positions. Appointments may be in the Department of Administrative Studies or in the Department of Economics within the Faculty or may be joint appointments in both departments depending on applicants' interests and experience. Salary Scale—Senior Lecturer: $A12,643-$A14,724 per annum with superannuation based on an endowment assurance scheme, the employee and employer contributing 5 percent and 10 percent respectively. Benefits: Travelling expenses for appointee and family; removal allowance, repatriation after three years' appointment if desired; temporary housing for an initial period. Study leave entitlement accumulates at the rate of one month's leave for each six month's service up to six years, with provision for financial assistance. Further general information and details of application procedure are available from the Academic Registrar, Monash University, Wellington Road, Clayton, Victoria 3168. Enquiries to the Dean, Professor D. Cochrane. Applicants should quote reference Nos. 20112M and 20112H. Closing Date: 15th October, 1974. The University reserves the right to make no appointment or to appoint by invitation.

Economics: Managerial economics and marketing at undergraduate and graduate level. Requirements include doctoral degree, teaching excellence and research potential. Three course load, with most courses meeting once weekly for two hours. Participation in theses committees expected. Twenty-two-man faculty in business administration and separate departments of economics, management science, statistics, etc. Any academic level can be considered. Send résumés and requests for information to Prof. Guy Black, Chairman, Business Administration, George Washington University, Washington, D.C. 20006 or call 202 676-6115.

Medical Economist: Joint appointment with the Department of Economics and the Division of Social Sciences and Humanities within the Health Sciences Center for an assistant or an exceptional associate professor. Primary course responsibilities: Introduction to Economics of Health (for medical students); Introduction to Economics and Welfare Economics (for social welfare students); Health Economics (offered through the Department of Economics). Send résumés to Professor John H. Wile, Department of Economics, State University of New York at Stony Brook, Stony Brook, New York 11794. Equal opportunity/affirmative action employer.

Economic Forecasting, Industrial Organization, Applied Microeconomics: The Graduate School of Business Administration of New York University has three faculty vacancies in the above-mentioned fields. Persons holding Ph.D.'s who are committed to quality teaching and excellence in scholarship are invited to apply. Applications are invited from women and persons with minority racial or ethnic backgrounds. Salaries are competitive. Please send résumé to Robert A. Kavesh, Chairman, Department of Economics, Graduate School of Business Administration, New York University, 90 Trinity Place, New York, New York 10006.

Economists Available for Positions

Italics indicate fields of specialization

Industrial Organization, International Trade and Monetary Theory, Economic Theory: Man, 30, married. B.A., M.A., Ph.D. completion date June 1974. Five years experience teaching principles and intermediate theory. Experience in university administration. Desires teaching position in U.S. or Canada. Résumé and references furnished on request. Available Fall 1974. E2581

Computer Assisted Instruction: Economist with interest and experience in computer assisted instruction in economics desires part-time, full-time, or consulting work in this area. E2589

Economics: Man, 25, married. B.A., M.A., in economics. Excellent grades, references. Desires position in economics in New York area. Strong background in microeconomics and familiar with mathematical analysis. Résumé will be furnished upon request. E2594

Consulting/Research, Public Policy Economics: Man, 31, Ph.D. from a major university. Three years working for the government of a major eastern city in a challenging, responsible position covering tax policy and citizen subsidies, also program planning budgeting analysis. Economist, major in economic development, dissertation on a proposed tax reform studied in an underdeveloped country, and minor in econometrics and international economics. Seeks a tough, challenging position for work on intricate public finance, regulatory and citizen subsidy policy. Prefers northeast or west coast with some opportunity for work abroad. Has a research proposal. E2596

Macroeconomics and Finance; Stabilization Policies, Economic Fluctuations and Forecasting, Money and Banking, International Finance, Financial Markets and Investments, Macro and Micro Principles: Man, 32, married, Ph.D. Seven years teaching and central banking experience. Desires government or teaching position which permits usage of experience in analyzing current business and financial conditions, both domestic and international, along with opportunities for some research. Especially interested in teaching undergraduates and economic education. Publication list and résumé furnished upon request. E2598

Advanced Microeconomics, Macroeconomics, Principles, Development-Planning-Regional—Urban Economics, Agricultural Economics, History of Economic Thought, Economic History and current problems of U.S.A., U.S.S.R., U.K., Japan, India. Comparative Economic Systems, Money, Banking and Public Finance: Woman economist, available for teaching, research and administrative position. Ph.D. London School of Economics. Two publications in top professional journals and two books to be published soon. Long teaching (graduate and undergraduate levels), research and administrative experience in America, Canada, England, and India. Currently associate professor of Mid-West State University. Will consider relocation as full professor. Résumé and references on request. E2601

International Economics, Economic Development, Microeconomics, Input-Output Economics, Labor Economics, Cost-effectiveness Studies, Economics of Natural Resources: Man, 46, Ph.D. 1963. Trilingual; several years of experience in teaching, government and business in Europe, Latin America, Canada, and U.S.A. Seeks teaching or research position. Ready to relocate. E2603

Managerial Economics, Environmental Economics, Price Theory: Man, 34, Ph.D. Nine years teaching experience. Teaching awards, publications, and scholastic honors. Desires position in Southeast in college administration, research, or government service. E2604

Economist-Administrator: Ph.D. (economics-finance & management) and J.D. (law). Male, 44, Married, U.S. nationality, business and academic (administration and faculty) experience, publications. American Men of Science; Who's Who in Consulting; Other. Currently, university administration. Desire administrative (Chairman, Dean, Vice-President, or President) or faculty professorship in applied economics-finance or management position. E2606

Economic Development Studies, Economic and Social Statistics, Industrial Studies, Economic Capacity, Labor and Business Administration, Applied Economics and Statistics: Man, 47, married. Considerable experience in government, academic, business, municipal corporations-statistical surveys, research, development planning, industrial-area-market research, feasibility studies, distribution aid surveys, techno-economic report, business management training and technical education for small scale industrialist-government. Extensive writing and publication experience. Participation in world conferences. Location open. Desire teaching and/or research and/or industrial planning-administrative position. Résumé and extensive personal evaluation credentials available immediately on request. E2607

International Economics and Trade, Microeconomics, Agricultural Economics, Development Economics, Middle East and South Asian Studies: Man, 38, married, Ph.D. in agricultural economics from a major university; fields in agricultural economics, international economics and development economics. Six years of research and teaching experience at B.Sc. and M.Sc. level with prominent Middle Eastern university. Applied research experience in South Asia, Near East and East Africa.

Desires teaching/research or teaching/administration position with college or medium sized university. References and résumé available on request. E2609

Economics Theory, Urban Economics, Managerial Economics, Economic Development, Consumer Economics, Labor Economics, Marketing Research: Male, 42, married, Ph.D. 1970. Past four years mostly administrative/research position, but teaching quarter time. Would like to teach and write. Extensive research background. Relates well to nonacademicians. Willing to participate in grant research if desirable. Prefer Southeast. E2610

Economic Theory, International Trade and Development, Finance and Planning and Public Administration: Man 34, Ph.D. in Economics in 1967 (West Germany). Doctoral research carried out in the pure theory of international economics. Publications, teaching and research experience. Desires teaching, research or academic position in U.S. or Canada. Available in 1974-1975. E2611

Urban & Regional Economics; Resource Economics; Land Use; Agricultural Economics: Man, 31, married, Ph.D. Three years experience in economic planning for central city in metropolitan area. Publications in economic base, housing market, and socio-demographic analyses and in capital budgeting. Served as projects leader. Desires position with corporation, federal or state government or quasi-public establishment. E2539

Urban and Regional, Micro., and Macro., Mathematical Economics, International and Development Economics, History of Econ. Thought: Male, 34, married. Ph.D. Several years teaching experience. Also economic planner with publications for a large city planning institution. Seeks teaching position in a college or university in U.S.A., Canada or England. Interested in innovative and quality education. Excellent references including student evaluation. Résumé on request. E2614

Micro and Macro Theory; Comparative Economic Systems; Soviet-type Economies; Economic Development and Planning; Interdisciplinary Social Science and Principles: Man, 32, married, Ph.D. (June 1974). Six years teaching experience (economics and international studies). NDEA Fellow, International Research and Exchanges Fellow (in Moscow, 1972-73). Desires college/university position in which dynamic teaching and relevant research are rewarded. Available September 1974. Résumé on request. E2615

Quantitative Economic Analysis, Systems Analysis, Cost Analysis and Modeling, Program Evaluation, Cost-Benefit Analysis: Man, 44, M.S. plus additional graduate study. Twenty years of experience in research and analysis of social/economic problems. Experience includes teaching, program planning and evaluation, cost-benefit analysis, economic analysis of aerospace, defense, welfare, medical, judicial, education, manpower development, state and local government planning, and other programs. Currently manager of an economic and cost analysis group of professionals in a non-profit corporation. Desire challenging managerial/analysis assignment. E2516

Macro Theory and Policy, Human Capital and Labor: Man, Ph.D., 40, married, experienced teacher (including graduate), publications, extensive faculty governance and committee experience. Desire full professorship with administrative possibilities, preferably not in Boston-to-Washington corridor. Excellent references. Available September 1975. Résumé upon request. E2617

Public Finance, Economic Development, Economic History, Economic Systems, Economic Theory, Business Courses: Man, 44, married; B.A. Honors, M.A., D.Sc. Eleven years of teaching experience; currently associate professor of economics; publications. Seeks teaching or research position. Available in September 1974. E2618

Public Finance, Urban Economics, Micro and Macroeconomics: Man, 39, Ph.D. Considerable teaching, research and administrative experience; publications. Desires research or teaching position with opportunity for research in the area of public finance, particularly state and local finance. Available July 1975. E2619

Development, Urban and Housing Economics, Econometrics: 57, D.Sc., Econ. Sciences. Available for senior position in the USA, Canada, or with international organizations, for university teaching and/or research. Background of 15 years teaching experience and about 20 years of research experience in the Central Government, United Nations work in the Far-East and Caribbean. Special eminence in organizing international seminars and training courses in the field of development planning, housing and urban economics, and the like. E2620

Consumer Economics, Forecasting, Quantitative Economics, Economic Theory: Man, 37, married; Ph.D. in quantitative economics. Has nine years experience in teaching economics and statistics. Now tenured at a large midwest university. Desires to move to a more challenging position. Publications in top journals. Enjoys teaching at both undergraduate and graduate levels. Available September 1974. Résumé on request. E2621

Macro, Micro, Money and Banking, Public Finance, Introductory Statistics, International Economics, History of Economic Thought, Principles, Economic Systems, Agricultural Economics, Market Structure and Asian Economics: Man 43, married, one child, B.S., LL.B., M.S. & Ph.D., 1962 and 1965, from leading land grant universities. 10 years undergraduate teaching and three years research experience. Presently tenured full professor in a Midwestern state college. Excellent

teacher with very good student evaluations. Modest publications, including a small book. Six years administrative experience. Wish to relocate; location open but prefers East, Mideast, Southeast, South and Midwest. Seeks College or University teaching and/or research position. Interested in undergraduate teaching, graduate program not essential. Would consider research position in private firms. References and résumé available on request. E2622

International Trade, Human Capital, Economic Development, Economic Theory: Man, 29, married, Ph.D., some publications. Teaching and research experience in the United States, Middle East, Africa. Consulting experience international organizations. Seeks teaching position West, Middle West, or Northeast. Available September 1975. Résumé on request. E2623

International Economics, Trade, International Finance, Economic Development: Man, 56, married, B.A. (Chicago), M.A. (Fletcher School of Law and Diplomacy), all work completed for Ph.D. except thesis. Has had almost 30 years of U.S. government experience, practically all with the Department of State and U.S. Foreign Service in the international economic field. Has held major supervisory and program direction positions such as Director of State Department's Office of International Trade and Finance; Minister for Economic and Commercial Affairs in American Embassy, Bonn; Minister for Political and Economic Affairs, American Embassy, New Delhi; Counselor for Economic Affairs, American Embassy, Yugoslavia; Deputy and Acting Director of U.S. Aid Mission, Yugoslavia; and Deputy Director of State Department's Research Bureau. Seeks senior executive position of operating or advisory character in international economic field with business or banking institution, private foundation or inter-governmental or other international agency. Available immediately. Résumé on request. E2624

Statistics (Including Sampling), Computer Applications Econometrics, Development Economics, National Accounts: Man, 30, married. B.A., M.A., Ph.D. (Economics). Five years teaching with emphasis on quantitative economics; four years as a statistician with the Statistical Office of the United Nations. Publications. Desires position in the New York Metropolitan Area. Résumé available on request. Available on reasonable notice. E2625

Economics, Finance, Cost Analysis: Female, 22, single. B.A. in economics with minors in political science and United States history. Plans to work on graduate degree in economics at night. Free to travel or relocate. Some experience as tax research assistant. Seeks a position as an economist or a job related to finance or cost analysis. References available upon request. E2626

Economics and Business: Man, Ph.D. Will take a one-year assignment. Dr. Clyde Crobaugh, 3300 S. Ocean Blvd., Delray Beach, Florida 33444.

Economic Gerontology, Natural Resource, Environmental, and Energy Economics, Micro-Macro Economics, Business and Government, and Industrial Organization: Man, 40, married Ph.D., dissertation in environmental economics; one year as government economist; last four years as assistant professor at midwest liberal arts college, interdisciplinary grants experience; seeks challenging position in teaching, research, and/or administration. Available June 1975. Résumé on request. E2627

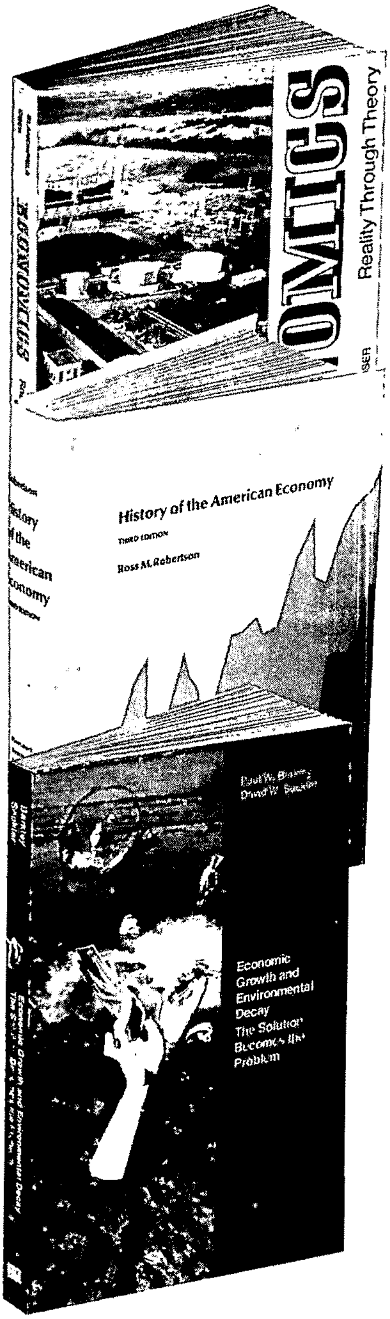
Economics-Business Administration: Male, 45, married, Ph.D. (International Finance), J.D., seeks position teaching some combination of business finance courses, business law courses, and international finance. Publications. Extensive teaching, administrative, law practice and business experience. E2628

Economics and Business Administration, Micro, Macro, Mathematical Economics, Econometrics and Statistics, Money and Banking, Public Finance, Managerial Finance, Managerial Economics, Quantitative Management, Economic Development, International Economics, Economic Systems: Ph.D. in economics, male, 41, 10 years of dynamic teaching experience and flexible at graduate and undergraduate levels. Currently an associate professor. Requests for résumé and interview are welcome. E2629

Microeconomic Theory, Labor Economics, Consumer Economics, History of Thought, Economic History: Man, 30, married; A.B., M.A., Ph.D. in dissertation stage. Seven years teaching and consulting experience; fellowships, scholarly papers and publications. Major text on consumer economics. Prefer upper midwest small college part or full time. E2630

Population/Family Planning (Economic Analysis, Surveys, Evaluation): Man, 38, married, M.A. (Economics); 7 years as junior economist with World Bank team and federal government in India; 3 years as statistician with international division of American organization in New York City. U.S. permanent resident. E2631

Economics for today's world



:CONOMICS
?eality Through Theory

WILLIAM C. BLANCHFIELD *and* JACOB OSER,
Jtica College of Syracuse University

3specially designed for the one-semester introductory course, his textbook provides a clear, lively account of the essential ıspects of macroeconomics, microeconomics, and major ecoıomic problems. Throughout the book, the authors relate theories o the student's own experiences and knowledge; emphasize ›ractical applications of the theories; evaluate the usefulness ›f economic theories as a practical tool, analyzing their strengths ınd weaknesses; and raise questions that go beyond the usual ange of problems found in standard economics textbooks. Each hapter concludes with a set of problems and essay questions ınd a brief bibliography. Accompanied by an Instructor's Aanual. Paperbound. 387 pages. $6.95

IISTORY OF THE AMERICAN ECONOMY
hird Edition

?OSS M. ROBERTSON, *Indiana University*

\ sound, readable account of the major events, institutions, and reative works that have shaped the American economic system, his highly successful textbook is thoroughly revised for this ediion to reflect current questions of public policy and to include he latest developments in economic history. Accompanied by a 'est Booklet with Lecture and Discussion Topics.
788 pages. $11.50

:CONOMIC GROWTH AND ENVIRONMENTAL DECAY
he Solution Becomes the Problem

'AUL W. BARKLEY, *Washington State University*
)AVID W. SECKLER, *Colorado State University*

irief, lively, and engaging, this book applies basic economic ıeories to current environmental problems. Throughout the ›xt, the authors focus on two major themes—the effect of ecoomic growth on society and the environment, and the relationıip between economics and the market system and the environıent—and they effectively demonstrate how an economic sys›m based on private decision-making processes can easily lead) environmental decay through gaps in ownership patterns, ıpses in incentives, and uncertainties regarding the future.
Paperbound. 193 pages. $3.95

HBJ Harcourt Brace Jovanovich, Inc. • New York • Chicago • San Francisco • Atlanta

Conglomerates Unlimited

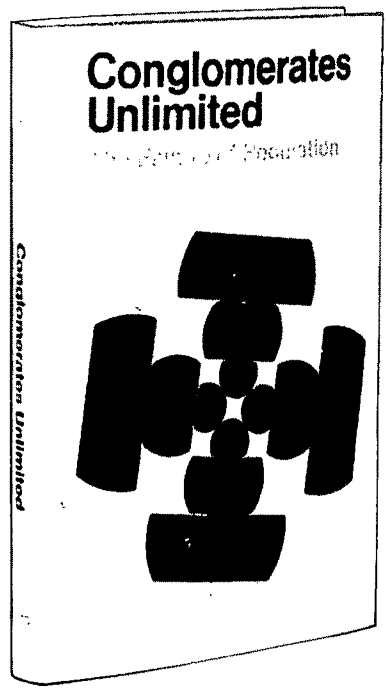
The Failure of Regulation

By John F. Winslow

Introduction by Irvin M. Grossack

This revealing book examines the methods and motives for building conglomerate corporations, the function of the central acquiring managements, and the role of government agencies in regulating these operations. Drawing from transcripts of congressional hearings, the author presents a detailed account of conglomerate building and its effect on our economy.

320 pages, index **$10.00**

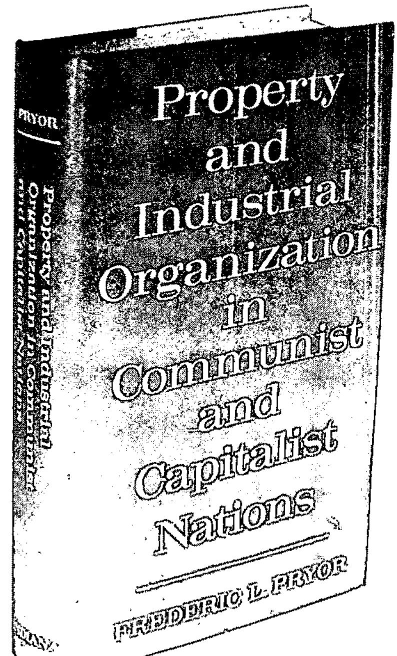


Property and Industrial Organization in Communist and Capitalist Nations

By Frederic L. Pryor

This major work presents the first comprehensive, systematic comparison of industrial organization among European and North American nations with different economic systems. It investigates patterns of nationalization, distribution of labor and property income, size distribution of industrial establishments and enterprises, degree of monopoly, spatial distribution of industry, and many other topics. (International Development Research Center, Studies in Development No. 7)

536 pages, index **$17.50**



INDIANA UNIVERSITY PRESS
10th & Morton Streets, Bloomington, Indiana 47401

The American Economic Review

ARTICLES

W. LEONTIEF Structure of the World Economy

E. J. KANE All for the Best: The Federal Reserve Board's *60th Annual Report*

J. E. STIGLITZ On the Irrelevance of Corporate Financial Policy

J. L. STEIN Unemployment, Inflation, and Monetarism

G. E. JOHNSON AND F. P. STAFFORD The Earnings and Promotion of Women Faculty

S. S. SHALIT AND U. BEN-ZION The Expected Impact of the Wage-Price Freeze on Relative Shares

G. W. SCULLY Pay and Performance in Major League Baseball

B. KLEIN Competitive Interest Payments on Bank Deposits and the Long-Run Demand for Money

D. LEVHARI AND Y. WEISS The Effect of Risk on the Investment in Human Capital

H. P. BINSWANGER The Measurement of Technical Change Biases with Many Factors of Production

J. MUELLBAUER Household Production Theory, Quality, and the "Hedonic Technique"

A. FELDMAN AND A. KIRMAN Fairness and Envy

SHORTER PAPERS: C. Henry; L. J. White; R. Cummings and V. Norton; A. E. Abrassart and D. D. McFarlane; A. C. Fisher, J. V, Krutilla, and C. J. Cicchetti; R. N. Batra and W. P. Russell; G. C. Rausser; D. M. G. Newbery; P. K. Bardhan and T. N. Srinivasan; C. C. Blaydon and P. W. Marshall; D. P. Baron; T. Russell; Y. Niho; P. L. Cheng.; N. Minabe.

DECEMBER 1974

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

• Published at George Banta Co., Inc., Menasha, Wisconsin.

• The American Economic Review, including four quarterly numbers, the *Proceedings* of the annual meetings, the *Directory*, and Supplements, is published by the American Economic Association and is sent to all members six times a year, in March, May, June, September, October, and December.

• Membership dues of the Association are $23.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature*. Junior membership at $11.50 per year is available to registered students for a maximum of three years. Subscriptions (libraries, institutions, or firms) are $34.50 a year, and only subscriptions to both publications will be accepted. Single copies of the *Review* and *Journal* are $5.00 each plus a $.50 per order service charge. Orders should be sent to the Secretary's office, Nashville, Tennessee.

• Correspondence relating to the *Papers and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address should be sent to the Secretary, Rendigs Fels, 1313 21st Avenue South, Nashville, Tennessee 37212. Change of address must reach the Secretary by the 1st of the month prior to the month of publication. The Association's publications are mailed second class and are not forwarded by the Post Office.

• Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

THE AMERICAN ECONOMIC REVIEW

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form and should be no longer than 50 pages of double-spaced typescript. A $10 submission fee must accompany each manuscript. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

December 1974

VOLUME LXIV, NUMBER 6

Articles

Structure of the World Economy
Wassily Leontief 823

All for the Best: The Federal Reserve Board's *60th Annual Report* *Edward J. Kane* 835

On the Irrelevance of Corporate Financial Policy
Joseph E. Stiglitz 851

Unemployment, Inflation, and Monetarism
Jerome L. Stein 867

The Earnings and Promotion of Women Faculty
George E. Johnson and Frank P. Stafford 888

The Expected Impact of the Wage-Price Freeze on Relative Shares
Sol S. Shalit and Uri Ben-Zion 904

Pay and Performance in Major League Baseball
Gerald W. Scully 915

Competitive Interest Payments on Bank Deposits and the Long-Run Demand for Money
Benjamin Klein 931

The Effect of Risk on the Investment in Human Capital *David Levhari and Yoram Weiss* 950

The Measurement of Technical Change Biases with Many Factors of Production
Hans P. Binswanger 964

Household Production Theory, Quality, and the "Hedonic Technique" *John Muellbauer* 977

Fairness and Envy
Allan Feldman and Alan Kirman 995

Shorter Papers

Investment Decisions Under Uncertainty: The "Irreversibility Effect" *Claude Henry* 1006
Industrial Organization and International Trade: Some Theoretical Considerations *Lawrence J. White* 1013
The Economics of Environmental Preservation:
Comment *Ronald Cummings and Virgil Norton* 1021
Comment *A. Eugene Abrassart and Dale D. McFarlane* 1025
Further Discussion *Anthony C. Fisher, John V. Krutilla, and Charles J. Cicchetti* 1030
Gains from Trade Under Uncertainty *Raveendra N. Batra and William R. Russell* 1040
Technological Change, Production, and Investment in Natural Resource Industries *Gordon C. Rausser* 1049
Cropsharing Tenancy in Agriculture:
Comment *D. M. G. Newbery* 1060
Rejoinder *P. K. Bardhan and T. N. Srinivasan* 1067
Incentive Contracts and Competitive Bidding:
Comment *Colin C. Blaydon and Paul W. Marshall* 1070
Reply *David P. Baron* 1072
Feige and Parkin on the Optimal Quantity of Money *Thomas Russell* 1074
Population Growth, Agricultural Capital, and the Development of a Dual Economy *Yoshio Niho* 1077
Comment on Corporate Income Taxes and the Cost of Capital: A Correction *Pao L. Cheng* 1086
Capital and Technology Movements and Economic Welfare *Nobuo Minabe* 1088

Errata 1101

In Memoriam: HAROLD HOTELLING 1102

Notes 1104

Titles of Doctoral Dissertations 1112

Structure of the World Economy

Outline of a Simple Input-Output Formulation

By WASSILY LEONTIEF*

I

The world economy, like the economy of a single country, can be visualized as a system of interdependent processes. Each process, be it the manufacture of steel, the education of youth, or the running of a family household, generates certain outputs and absorbs a specific combination of inputs. Direct interdependence between two processes arises whenever the output of one becomes an input of the other: coal, the output of the coal mining industry, is an input of the electric power generating sector. The chemical industry uses coal not only directly as a raw material but also indirectly in the form of electrical power. A network of such links constitutes a system of elements which depend upon each other directly, indirectly, or both.

The state of a particular economic system can be conveniently described in the form of a two-way input-output table showing the flows of goods and services among its different sectors, and to and from processes or entities ("value-added" and "final demand") viewed as falling outside the conventional borders of an input-output system. As the scope of the inquiry expands, new rows and columns are added to the table and some of the external inflows and outflows become internalized. Increasing the number of rows and columns that describe an economic system also permits a more detailed description of economic activities commonly described in highly aggregative terms.

Major efforts are presently underway to construct a data base for a systematic input-output study not of a single national economy but of the world economy viewed as a system composed of many interrelated parts. This global study, as described in the official document, is aimed at

> . . . helping Member States of the United Nations make their 1975 review of world progress in accelerating development and attacking mass poverty and unemployment. First, by studying the results that prospective environmental issues and policies would probably have for world development in the absence of changes in national and international development policies, and secondly, by studying the effects of possible alternative policies to promote development while at the same time preserving and improving the environment. By thus indicating alternative future paths which the world economy might follow, the study would help the world community to make decisions regarding future development and environmental policies in as rational a manner as possible.[1]

Preliminary plans provide for a description of the world economy in terms of twenty-eight groups of countries, with

* Harvard University. This article is the lecture Wassily Leontief delivered in Stockholm, Sweden, December 1973, when he received the Nobel Prize in Economic Science.

The author is indebted to Peter Petri for setting up and performing all the computations, the results of which are presented in this lecture, and to D. Terry Jenkins for preparing the graphs and editorial assistance.

[1] Quoted from: "Brief Outline of the United Nations Study on the Impact of Prospective Environmental Issues and Policies on the International Development Strategy," Apr. 1973.

about forty-five productive sectors for each group. Environmental conditions will be described in terms of thirty principal pollutants; the use of nonagricultural natural resources in terms of some forty different minerals and fuels.

DEVELOPED COUNTRIES

	Extraction Industry	Other Production	Abatement Industry	FINAL DEMAND Domestic	FINAL DEMAND Trade	Total Output
Extraction Industry	0	76	0	2	-15	63
Other Production	21	1809	21	2414	19	4284
Pollution	5	62	-63	60	0	64
Employment	18	1372	20	287	0	
Other Value Added	21	996	22	0	0	

LESS DEVELOPED COUNTRIES

	Extraction Industry	Other Production	Abatement Industry	FINAL DEMAND Domestic	FINAL DEMAND Trade	Total Output
Extraction Industry	0	8	0	2	15	25
Other Production	7	197	0	388	-19	573
Pollution	2	8	0	11	0	21
Employment	9	149	0	99	0	
Other Value Added	8	220	0	0	0	

FIGURE 1. WORLD ECONOMY IN 1970
(Billions of 1970 dollars)

II

The subject of this lecture is the elucidation of a particular input-output view of the world economy. This formulation should provide a framework for assembling and organizing the mass of factual data needed to describe the world economy. Such a system is essential for a concrete understanding of the structure of the world economy as well as for a systematic mapping of the alternative paths along which it could move in the future.

Let us consider a world economy consisting of 1) a Developed and 2) a Less Developed Region. Let us further divide the economy of each region into three productive sectors: an Extraction Industry producing raw materials; All Other Production, supplying conventional goods and services; and a Pollution Abatement Industry. In addition to these three sectors, there is also a consumption sector specified for each region. The function of the Abatement Industry is to eliminate pollutants generated by the productive sectors, consumers, and the Abatement Industry itself.

The two input-output tables displayed as Figure 1 describe the intersectoral flows of goods and services within the Developed and the Less Developed Economies. The flow of natural resources from the Less Developed to the Developed Countries, as well as the opposite flow of Other Goods from the Developed to the Less Developed Countries are entered in both tables: positively for the exporting region, and negatively for the importing region.

In each of the two tables the right-most entries in the first and second row represent the total domestic outputs of the Extraction Industry and of Other Production, respectively.

Each positive number along the third (pollution) row shows the physical amount of pollutant generated by the activity named at the head of the column in which that number appears. The negative quantity shown at the intersection of the third column and the third row represents the amount of pollutant eliminated by Abatement activities.[2] Inputs such as power, chemicals, etc. purchased by the Abatement Industry from other sectors, and value-added paid out by that industry are entered as positive amounts in the same third column. The difference between the total amount of pollution generated in all sectors and the amount eliminated by the Abatement sector is represented by the *net* emission figure, the right-most entry in the third row. Finally, labor inputs used in each sector and payments made to other

[2] All quantities are measured in billions of dollars "in current prices;" pollutants are "priced" in terms of average "per unit" abatement costs.

income-receiving agents are shown in the bottom two rows.

The numbers in these two tables are, strictly speaking, fictitious. But their general order of magnitude reflects crude, preliminary estimates of intersectoral flows within and between the Developed and Less Developed Regions during the past decade.

For analytical purposes, the outputs and inputs of the Extraction Industry and Other Production, as well as the amounts of pollutants generated and abated, can be interpreted as quantities measured in the appropriate physical units (pounds, yards, kilowatts, etc.). The same is true of the services of some of the so-called primary factors: labor inputs, for example, are entered in the second to last row of each table. A similar physical measurement of the other components of value-added, even if it were possible in principle, is impossible given the present state of knowledge. In pure, or should I say speculative, economic theory, we can overcome this kind of difficulty by introducing some convenient albeit unrealistic assumptions. But a theoretical formulation designed to permit empirical analysis has to account for the fact that at least some components of value-added cannot be interpreted as payments for measurable physical inputs, but must be treated as purely monetary magnitudes.

III

The flows described in the two input-output tables are interdependent. They have to satisfy three distinct sets of constraints. First, within each production or consumption process there exists a technological relationship between the level of output and the required quantities of various inputs. For example, if we divide each figure in the first column of the first table of Figure 1 (the inputs of the Extraction Industry) by the total output of that sector (the last figure in the first row), we find that to produce one unit of its output this sector absorbed .3372 units of the output of Other Production, used .2867 units of Labor Services, and spent .3332 dollars for other value-added. Moreover, for each unit of useful output the Extraction Industries generated .0859 units of pollution. Other sets of input-output coefficients describe the technical structure of every sector of production and consumption in both groups of countries.

While statistical input-output tables continue to serve as the principal source of information on the input requirements or "cooking recipes" of various industries, increasingly we find economists using engineering data as a supplemental source. Complete structural matrices of the two groups of countries used in our example are shown in Figure 2.

The second set of constraints that has to be satisfied by every viable system requires that the total (physical) amounts of outputs and inputs of each type of good must be in balance, i.e., total supply must equal total demand. In the case of a pollutant, *net* emission must equal the total

Technical and Consumption Coefficients

Developed Countries

$$A_1 = \begin{bmatrix} .0 & .0178 & .0 \\ .3372 & .4223 & .3298 \\ .0859 & .0144 & .0118 \end{bmatrix} \qquad C_1 = \begin{bmatrix} .0007 \\ .8834 \\ .0218 \end{bmatrix}$$

$$l_1 = [.2867 \quad .3203 \quad .3161] \qquad l_1^c = [.1050]$$

$$r_1 = [.3332 \quad .2324 \quad .3482] \qquad r_1 = [.0 \quad]$$

Less Developed Countries

$$A_2 = \begin{bmatrix} .0 & .0141 & .0 \\ .2934 & .3437 & .3298 \\ .0859 & .0144 & .0118 \end{bmatrix} \qquad C_2 = \begin{bmatrix} .0037 \\ .7943 \\ .0218 \end{bmatrix}$$

$$l_2 = [.3729 \quad .2597 \quad .3161] \qquad l_2^c = [.2020]$$

$$r_2 = [.3337 \quad .3825 \quad .3541] \qquad r_2^c = [.0 \quad]$$

FIGURE 2

The coefficients in these tables do not sum to unity because the pollution generated by industry and by final demand is only partially abated in the developed countries and not abated at all in the less developed countries.

amount generated by all sectors less the amount eliminated by the abatement process.

For example, the balance between the total output and the combined inputs of extracted raw materials can be described by the following equation:

$$\underbrace{(1-a_{11})x_1}_{\text{net output of Extraction Industry}} \quad \underbrace{-a_{12}x_2}_{\text{amount delivered to Other Production}} \quad \underbrace{-a_{13}x_3}_{\text{amount delivered to the Abatement Industry}} \quad \underbrace{-c_1y}_{\text{amount delivered to Final Users}} \quad \underbrace{-T_1}_{\text{amount exported}} = 0 \tag{1}$$

The equation describing the balance between generation, abatement and net emission of pollution reads as follows:

$$\underbrace{-a_{31}x_1-a_{32}x_2}_{\text{gross amount of pollution generated by sectors 1 and 2}} \quad \underbrace{+(1-a_{33})x_3}_{\text{amount abated by abatement activities}} \quad \underbrace{-c_3y}_{\text{gross amount generated by consumers and government}} \quad \underbrace{+E}_{\text{net amount emitted into the environment}} = 0 \tag{2}$$

x_1 and x_2 represent the total outputs of the Extraction Industry and of Other Production, respectively; x_3, the level of activity of the abatement sector; y, the sum total of values-added, i.e., Gross National Income. The "technical coefficient" a_{ij} represents the number of units of the product of sector i absorbed (or generated in the case of pollution) by sector j in producing one unit of its output; c_j is a "consumption coefficient" describing the number of units of the output of sector j consumed (or generated in the case of pollution) per unit of total value-added, i.e., per unit of Gross National Income.

Figure 3 displays the complete set of linear equations describing the physical balances between outputs and inputs of all sectors in both countries in terms of compact matrix notation. The last of these equations—written below in its explicit form—describes the flows of exports and imports that link the Developed and Less Developed areas into a single world economy.

$$B = T_2p_2 - T_1p_1 \tag{3}$$

The balance of trade B, i.e., the difference between the monetary value of the two opposite trade flows, depends not only on the quantities T_1 and T_2 of traded goods but also on their prices, p_1 and p_2. The higher the price a country receives for its exports, or the lower the price it pays for imports, the better are its "terms of trade."

The last of the three sets of relationships describes the interdependence of the prices of all goods and services and the values-added paid out, per unit of output, by each industry. For example, a typical equation in this set states that the price at which the Extraction sector sells one unit of its output equals the average outlay incurred in producing it. This includes the costs (i.e., quantities x prices) of inputs purchased from other sectors, wages paid and all other value-added:

$$\underbrace{p_1}_{\text{price of output}} \quad \underbrace{-a_{11}p_1-a_{21}p_2}_{\text{cost of material inputs}} \quad \underbrace{-q_1a_{31}p_3}_{\text{cost of pollution abatement}} \quad \underbrace{-l_1w}_{\text{cost of labor inputs}} \quad \underbrace{-r_1}_{\text{other value-added}} = 0 \tag{4}$$

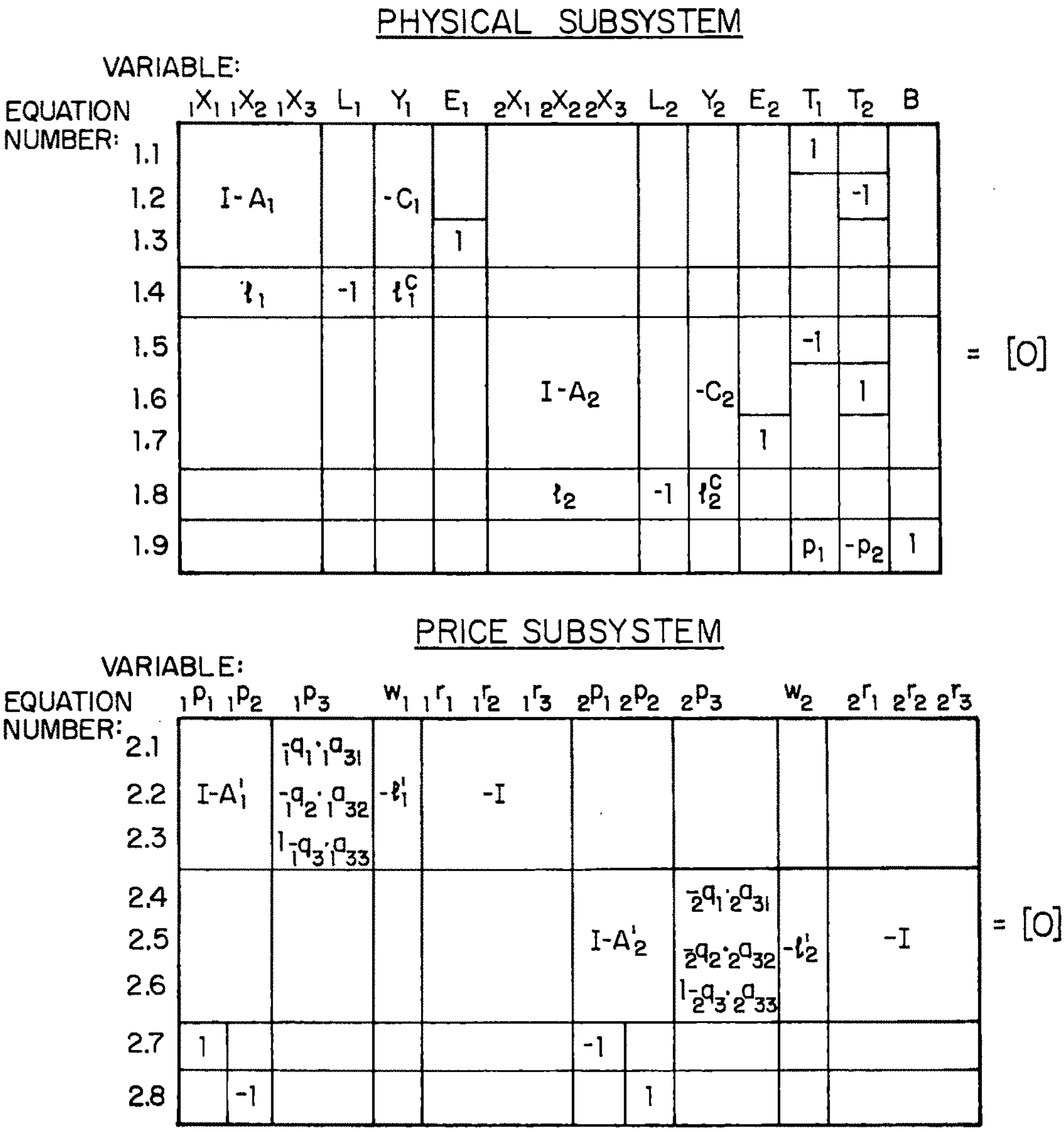


FIGURE 3

The technical coefficients (a_{ij} and l_i) appearing in this equation are the same as those appearing in the structural matrices of Figure 2. The abatement ratios q_i represent the fraction of the gross pollution emission of industry i that is eliminated (at that industry's expense)[3] by the Abatement Industry.

In this example, the system of physical balances contains 9 equations with 15 variables, while the price-values-added system has 8 equations with 14 variables. But these 14 variables are reduced to 12 and the number of equations to 6 if one assumes

[3] This formulation is based on the assumption that the pollution generated by a particular sector is being eliminated at its own expense. In case the abatement cost is being paid out by the government out of its tax revenues, the price equations have to be modified accordingly. See the author and Milton Moss.

from the outset that the internationally traded products of the Extraction Industry and Other Production have the same price in the Developed and the Less Developed Countries. Equations 2.7 and 2.8 worked out explicitly read:

$$(5) \quad {}_1p_1 = {}_2p_1(\equiv p_1) \quad \text{and} \quad {}_1p_2 = {}_2p_2(\equiv p_2)$$

The combination of both systems viewed as a whole contains 29 unknowns but only 17 equations. Thus, to arrive at a unique solution, we have to fix the values of 12 variables on the basis of some outside information, i.e., their values have to be determined exogenously.

Two types of quantitative information are required for the solution of this system. First, some data are used in the form of appropriate structural coefficients. Other kinds of factual information are introduced by assigning specific numerical values to appropriate "exogenous" variables.

In view of the uneven quality of data that will constitute the empirical basis of the present inquiry, it would be a tactical mistake to pour all the factual information we possess into the rigid mold of a single, all-embracing, inflexible explanatory scheme. The decision of which variables should be treated as dependent and which should be fixed exogenously is essentially a tactical one. The theoretical formulation is a weapon; in deciding how to use it we must take into account the nature of the particular empirical terrain.

To assess the influence of factors considered external to our theoretical description of the world economy, we earmark 6 physical and 5 value-added variables as "exogenous." Figures 4 and 5 show which variables are endogenous and assign values to all exogenous variables. These assumptions permit us to project changes in our simple world economy from a state representative of the present ("1970") to three alternative hypothetical states about thirty years hence ("2000 (I)," "2000 (II)," and "2000 (III)").

Total labor input in Developed Countries, L_1, is exogenous: under full or nearly full employment, its magnitude depends on demographic and cultural factors not accounted for within our formal theoretical system. Substantial endemic unemployment in the Less Developed Countries makes it advisable to consider the level of total employment as depending on the level of output—that is, to treat L_2 as endogenous.

The output of the Extraction Industry in the Developed Countries is restricted by the limited availability of natural resources. We account for this limitation by making ${}_1x_1$ exogenous. In the Less Developed Countries, where natural resources are still plentiful, the output of the Extraction Industry, ${}_2x_1$, depends partly on a small domestic market but primarily on the import requirements of Developed Countries. Thus, ${}_2x_1$ can be treated as a dependent variable.

The situation is reversed in the case of Other Production. In Developed Countries the output of manufactured goods normally adjusts to the level of final demand, making ${}_1x_2$ a dependent variable. Yet in the Less Developed Countries the output of Other Production, ${}_1x_2$, is restricted by external factors such as weak infrastructure and limited capital. In this case rising domestic inputs usually stimulate a growing demand for imports. Hence, ${}_2x_2$ is treated as independent and T_1 and T_2 as dependent variables.

In the price-value-added system of equations, all money wages and other value-added payments in the Developed Countries (w, r_1, r_2, and r_3) are exogenously determined. This means that the prices of all three products can be derived endogenously. In Less Developed Countries the situation seems to be different: since the

VARIABLES		DEVELOPED COUNTRIES Case I	DEVELOPED COUNTRIES Case II	DEVELOPED COUNTRIES Case III	LESS DEVELOPED COUNTRIES Case I	LESS DEVELOPED COUNTRIES Case II	LESS DEVELOPED COUNTRIES Case III
Extraction output	X_1	Capacity limited to 150% of 1970 levels			Endogenous		
Other production	X_2	Endogenous			Capacity grows 6.4% per annum between 1970 and 2000		
Abatement output	X_3	Endogenous			0	Endogenous	
Employment	L	Increase proportional to population increase			Endogenous		
Final Demand	Y	Endogenous					
Net pollution emission	E	Limited to current levels assuming 1970 standards			Endogenous	Limited to twice 1970 levels	
Net trade in Extractive goods	T_1	Endogenous					
Net trade in Other goods	T_2						
Trade balance	B	A deficit for Less Developed Countries amounting to 1% of Developed Countries' income, reflecting capital flows and aid					
Technical Coefficients	A	Unchanged from 1970		Twice 1970 levels for Extraction Industry	Unchanged from 1970		
Labor Coefficients	l	1/3 1970 levels, due to increased productivity		2/3 1970 levels for Extraction Industry	1/3 1970 levels due to increased productivity		
Consumption coefficients	C	Unchanged from 1970					
Extraction goods price	P_1	Obtained from solution of price system					
Other goods price	P_2						

FIGURE 4. PHYSICAL SYSTEM ASSUMPTIONS

prices of commodities produced by Extraction and Other Production are determined by the cost of their production (including the exogenous value-added) in the Developed Countries, the value-added that can be paid out by the two sectors producing these goods in the Less Developed Countries, ${}_2r_1$ and ${}_2r_2$, simply reflect the difference between a given price and the production costs.

Raw materials are, as a rule, relatively more abundant and more cheaply extracted in Less Developed Countries, thus the value-added earned by Extraction Industries in Less Developed Countries can be expected to be relatively high. Ricardo speaks in this connection of "mining rents." On the other hand, technical input coefficients or, more properly, costs in Other Production of the Less Developed

<table>
<tr><th colspan="2" rowspan="2">VARIABLES</th><th colspan="3">DEVELOPED COUNTRIES</th><th colspan="3">LESS DEVELOPED COUNTRIES</th></tr>
<tr><th>Case I</th><th>Case II</th><th>Case III</th><th>Case I</th><th>Case II</th><th>Case III</th></tr>
<tr><td>Extraction goods price</td><td>P_1</td><td colspan="6" rowspan="3">Endogenous</td></tr>
<tr><td>Other goods price</td><td>P_2</td></tr>
<tr><td>Abatement Price</td><td>P_3</td></tr>
<tr><td>Wage rate</td><td>w</td><td colspan="6">Kept at 1970 level (index=1.0)</td></tr>
<tr><td>Other value-added in Extraction per unit of output</td><td>r_1</td><td colspan="3" rowspan="3">Kept at 1970 levels (index=1.0)</td><td colspan="3" rowspan="2">Endogenous</td></tr>
<tr><td>Other value-added in Other Production</td><td>r_2</td></tr>
<tr><td>Other value-added in Abatement</td><td>r_3</td><td colspan="3">Kept at 1970 level (index=1.0)</td></tr>
<tr><td>Technical coefficients</td><td>A</td><td colspan="2">Unchanged From 1970</td><td>Twice 1970 levels for Extraction Industry</td><td colspan="3">Unchanged from 1970</td></tr>
<tr><td>Labor coefficients</td><td>l</td><td colspan="2">1/3 1970 levels, due to increased productivity</td><td>2/3 1970 levels for Extraction Industry</td><td colspan="3">1/3 1970 levels, due to increased productivity</td></tr>
<tr><td>Abatement coefficients</td><td>q</td><td colspan="6">$q_1=q_2=q_3=x_3/(x_3+E)$, that is, all Abatement coefficients of a given country are set to a value that reduces net pollution to the exogenously specified level E</td></tr>
</table>

FIGURE 5. PRICE SYSTEM ASSUMPTIONS

Countries can be expected to be higher than in Developed Countries. Because of this, the value-added earned per unit of output in that sector tends to be relatively low.

Since a principal purpose of the aforementioned United Nations project is a "realistic evaluation of the effects of alternative types of environmental policies on the economic prospects of Less Developed Countries," net pollution emissions E_1 and E_2 are treated as exogenously determined in two of our projections.

Assigning specific numerical magnitudes to all exogenously determined variables permits effective use of a variety of external data in arriving at a unique numerical solution of the formal input-output system. As the empirical inquiry advances, exogenous variables can be internalized through introduction of additional equations.

The most important but also the most demanding step in implementing an empirical input-output system is the determination of values of hundreds or even thousands of structural coefficients. The relevant methodologies are so varied and specialized that I abstain from discussing them in this general context.

IV

As has been explained above, three different sets of factual assumptions provided

the basis for the three alternative projections of the state of one simple world economy for the year "1970" to the year "2000." Figures 4 and 5 contain their full specification, while the results of the computations are summarized in three pairs of input-output tables presented in the Appendix.

The bar charts displayed in Figures 6 and 7 facilitate a systematic examination of these findings. The width of each bar represents the relative size of the corresponding economic activity measured in base-year dollars. The length of each bar indicates the percentage increase or decrease in the level of each activity as the world economy passes from one state to another. Exogenous variables are identified by asterisks.

The long bars in the uppermost rows of these economic profiles indicate an upsurge in output and total consumption and a downward movement of prices: a "great leap forward" from 1970 to 2000. Case I is a projection that critically depends on two assumptions. First, the employed labor force in Developed Countries will in-

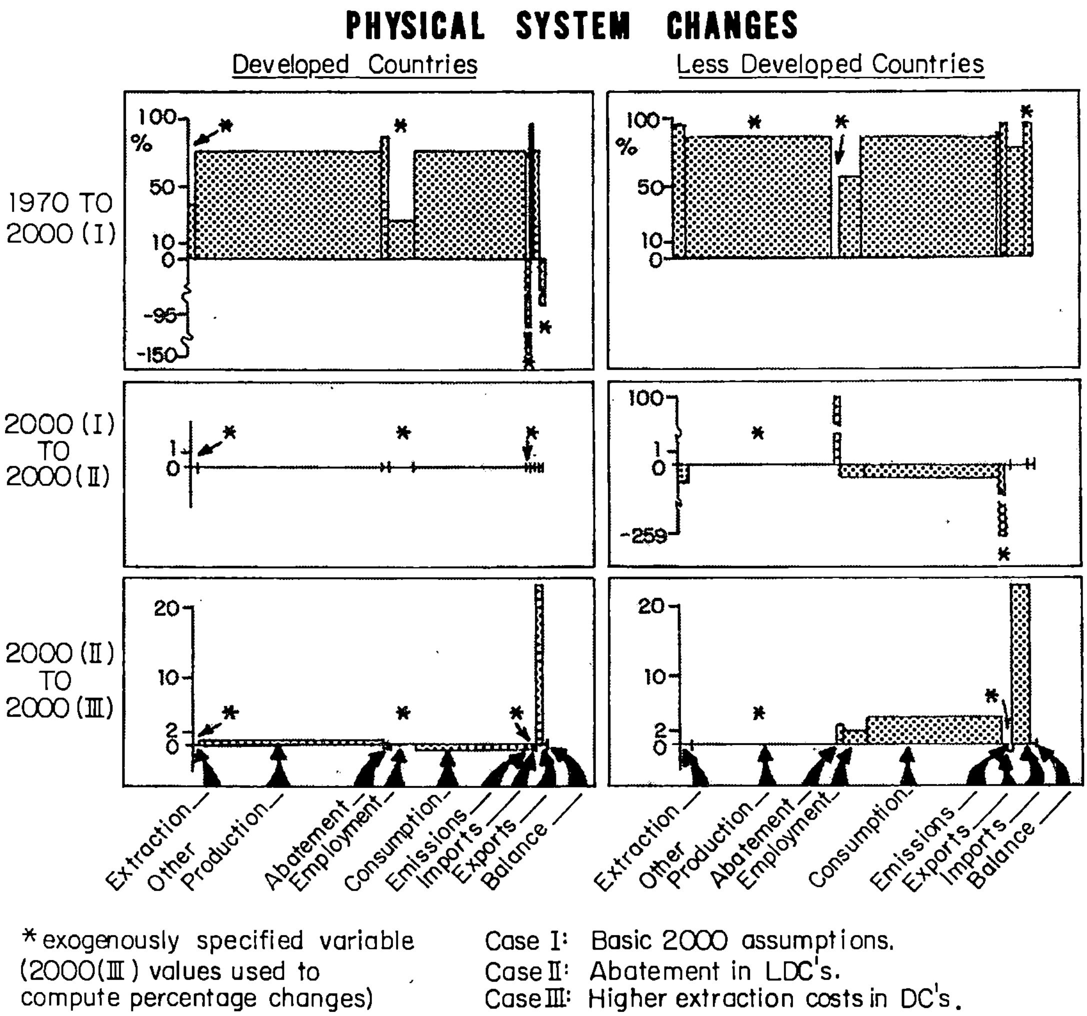


FIGURE 6. PHYSICAL SYSTEM CHANGES

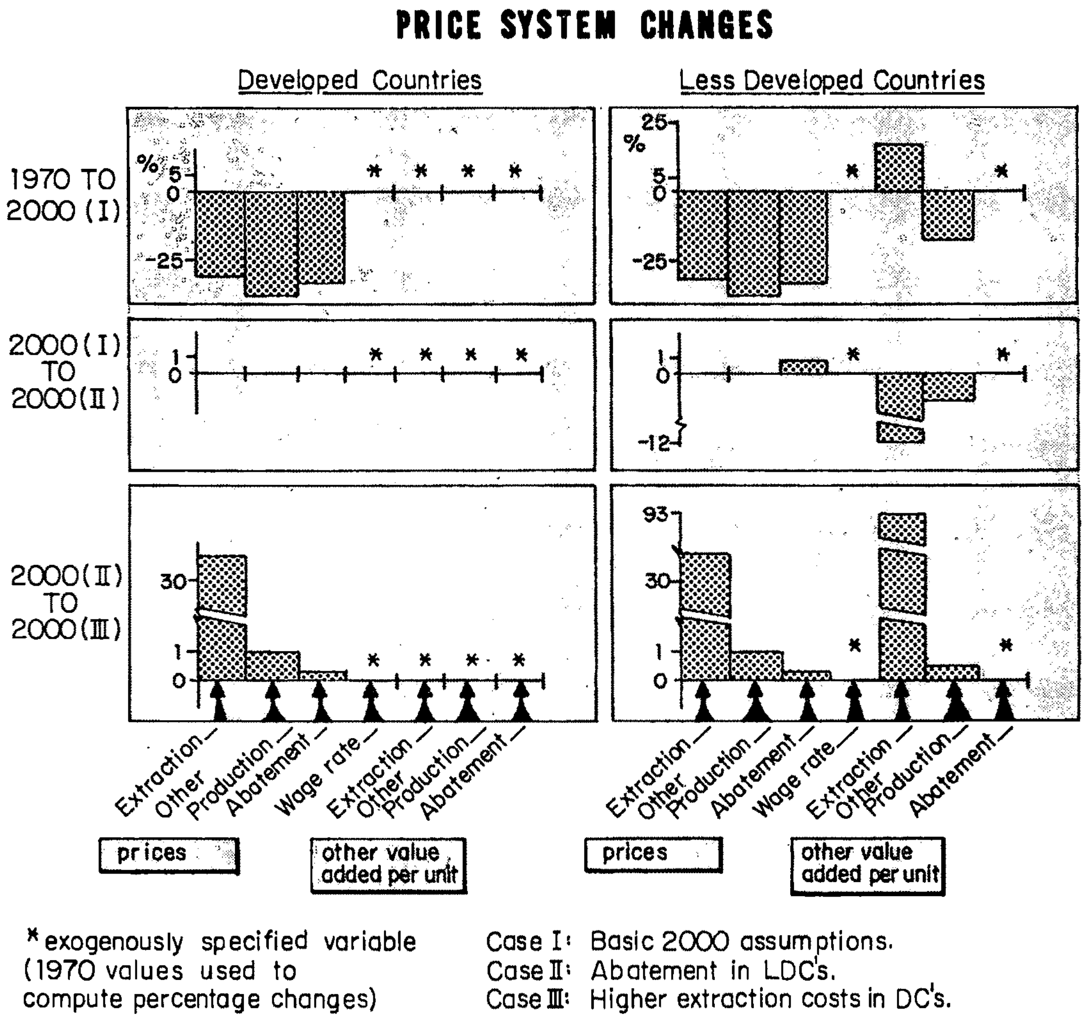


FIGURE 7

crease with population growth. Second, labor productivity in both regions (the reciprocal of the labor coefficient) will be three times as high in 2000 as in 1970, with all other input coefficients remaining the same. Strict enforcement of standards contained in the United States Clean Air Act of 1967 (as amended in 1970) will bring about a sharp drop in unabated emissions in the Developed areas, while in Less Developed Countries the absence of any abatement activity will force the pollution level up. International trade will expand faster than domestic economic activities. Prices (measured in wage units) will decline, while the value-added in Less Developed Countries will rise in the Extraction Industry but fall in Other Production.

How would the future economic picture change if strict antipollution standards were also observed in Less Developed Countries? The answer is presented in the second row of bar graphs on Figure 6 and 7. In the Developed Countries there will be practically no change. In Less Developed Countries the inauguration of abatement activities aimed at limiting pollution to

twice its 1970 level would bring about expanded employment while requiring some sacrifices in consumption. Value-added would fall sharply in the Extraction Industry and somewhat less in Other Production.

How would the situation thus attained be affected by a significant increase in the operating costs of the Extraction Industry in the Developed Countries? The bottom row of profiles in Figures 6 and 7 shows how the conditions in both regions of the world economy would be affected if the productivity of labor in the Extraction Industry of Developed Countries rose only $1\frac{1}{2}$ rather than 3 times between 1970 and 2000 while the amounts of other Extraction inputs doubled per unit of output. The output of Other Production in the Developed Countries would register a slight increase and the level of consumption a slight decrease. Consumption in the Less Developed Countries would experience a substantial increase. The mechanism responsible for such a redistribution of income between the Developed and Less Developed Countries involves a steep increase in the price of Extraction goods compared to other prices, a corresponding rise in value-added (rents yielded by the Extraction Industry of the Less Developed Countries), and, finally, a substantial increase in imports accompanied by slight reduction of exports from these countries, both reflecting a marked improvement in their "terms of trade."

I refrain from drawing any factual conclusion from the economic projections presented above. The computer received fictitious inputs and necessarily issued fictitious outputs. All theories tend to shape the facts they try to explain; any theory may thus turn into a procrustean bed. Our proposed theoretical formulation is designed to protect the investigator from this danger: it does not permit him to draw any special or general conclusions before he or someone else completes the always difficult and seldom glamorous task of ascertaining the necessary facts.

Appendix

Projected World Economy in 2000 (Case I)
(Billions of 1970 dollars)

DEVELOPED COUNTRIES

	Extraction Industry	Other Production	Abatement Industry	FINAL DEMAND Domestic	FINAL DEMAND Trade	Total Output
Extraction Industry	0	316	0	8	-226	98
Other Production	33	7502	160	9713	357	17765
Pollution	8	256	-479	240	0	25
Employment	9	1897	51	379	0	
Other Value-Added	33	4129	169	0	0	

LESS DEVELOPED COUNTRIES

	Extraction Industry	Other Production	Abatement Industry	FINAL DEMAND Domestic	FINAL DEMAND Trade	Total Output
Extraction Industry	0	52	0	12	226	290
Other Production	85	1255	0	2668	-357	3650
Pollution	25	53	0	73	0	151
Employment	36	316	0	226	0	
Other Value-Added	112	1143	0	0	0	

Projected World Economy in 2000 (Case II)
(Billions of 1970 dollars)

DEVELOPED COUNTRIES

	Extraction Industry	Other Production	Abatement Industry	FINAL DEMAND Domestic	FINAL DEMAND Trade	Total Output
Extraction Industry	0	316	0	8	-226	98
Other Production	33	7502	160	9713	357	17765
Pollution	8	256	-479	240	0	25
Employment	9	1897	51	379	0	
Other Value Added	33	4129	169	0	0	

LESS DEVELOPED COUNTRIES

	Extraction Industry	Other Production	Abatement Industry	FINAL DEMAND Domestic	FINAL DEMAND Trade	Total Output
Extraction Industry	0	52	0	12	226	290
Other Production	85	1254	36	2632	-357	3650
Pollution	25	53	-108	72	0	42
Employment	36	316	12	223	0	
Other Value Added	100	1118	39	0	0	

PROJECTED WORLD ECONOMY IN 2000 (CASE III)
(Billions of 1970 dollars)

DEVELOPED COUNTRIES

	Extraction Industry	Other Production	Abatement Industry	FINAL DEMAND		Total Output
				Domestic	Trade	
Extraction Industry	0	315	0	8	-225	98
Other Production	66	7472	159	9678	461	17836
Pollution	8	255	-477	239	0	25
Employ-ment	19	1890	51	378	0	
Other Value-Added	33	4112	168	0	0	

LESS DEVELOPED COUNTRIES

	Extraction Industry	Other Production	Abatement Industry	FINAL DEMAND		Total Output
				Domestic	Trade	
Extraction Industry	0	51	0	13	225	289
Other Production	85	1254	37	2735	-461	3650
Pollution	25	53	-111	75	0	42
Employ-ment	36	316	12	232	0	
Other Value Added	189	1125	40	0	0	

REFERENCES

W. Leontief, "Environmental Repercussions and the Economic Structure—An Input-Output Approach," *Rev. Econ. Statist.*, Aug. 1970, *52*, 262–71.

M. Moss, *The Measurement of Social and Economic Performance*, Nat. Bur. Econ. Res. *Stud. in Income and Wealth*, Vol. 38, New York 1973.

All for the Best: The Federal Reserve Board's *60th Annual Report*

By EDWARD J. KANE*

Economists don't just discuss Federal Reserve (*FR*) documents. We virtually *pounce* on them. We criticize everything from the frailest economic argument to the color scheme and quality of the binding. Except for newsy details necessary to keep our preachments up to date, the issues we raise seldom change. Decade after decade, we upbraid the Fed about the responsibilities for economic statesmanship that accompany its statutory independence; its unwillingness to specify an explicit model of how it believes that its policies impact on economic variables; its special concern for cushioning its effects on the money markets; the inevitable ineptness of its interventions into specific markets; and its extraordinary penchant for humbug.

Fed officials must long have wondered that we don't tire of the game. While professional rewards for scolding the Fed are not inconsiderable, a zealous desire to uphold professional standards seems to be the driving force. The characteristic pragmatism and electicism espoused by obviously well-intentioned Fed officials make them appear as willing students, anxious to have their misconceptions replaced by true knowledge. Also, we feel a compulsion to speak out against the roseate and self-serving analysis imbedded in typical Fed pronouncements. While none of this earns us marks for political sophistication, it fosters our professional self-image and keeps alive the ideal of an uncompromising central bank maximizing (as far as it can on its own) national economic welfare.

More than any other regular *FR* publication, the Board's *Annual Report* reminds us that the Fed operates in an intensely political environment and devotes considerable energy to maintaining a wishy-washy political stance. Submitted to the Speaker of the House, the *Report* is a ritual act of obeisance designed to inform congressional overlords of what and how well the system has performed for them during the preceding year. The *Report* makes no attempt to educate the public about the critical policy issues over which different political constituencies are fighting. Nor does it take on other constituencies to mobilize support for the Fed's own reading of the nation's policy needs. Far from building a constituency for controversial Fed policies, the *Report* is used to soften conflict and to mend fences. The *Report*'s goal is to put the best face possible on Fed actions of the past year. No uncontroversial central-banking chore is too small to be reported; no controversial action is large enough to be reported in proper political perspective. Except to someone trained in the art of decoding Fed statements, the *FR* Board's sixty *Annual Reports* provide little support for the hypothesis that a guilty conscience needs no accuser. Fed policies (and the motives behind them) are often very different from what the Fed seems to say they are.

* Professor of banking and monetary economics, Ohio State University. This is a review article of the *60th Annual Report, 1973* of the Board of Governors of the Federal Reserve System, Washington, May 31, 1974. The author offers thanks to the National Science Foundation for financial support, to Thomas Havrilesky, William Poole, Steven Weiss, Stephen Zeller, and the managing editor of this *Review* for helpful criticisms of an earlier draft, and to Stephen Zeller a second time for compiling and processing needed data.

Formally, each *Report* is divided into two parts: a brief narrative account of "Monetary Policy and the U.S. Economy in 19xx" and a comprehensive collection of statements and figures covering System "Records, Operations, and Organization." In the 1973 *Report*, Part 2 is four times the size of Part 1.

The longest and most analytical sections of the *Report* stand at the beginning of Part 2. These are the official records of policy actions taken by the *FR* Board, pp. 73–127, and by the Federal Open Market Committee (*FOMC*), pp. 128–223. Very short sections report on: operations in foreign currencies (2 pp.); the Voluntary Foreign Credit Restraint Program (4 pp.); legislation enacted (3 pp.); legislative recommendations (8 pp.); litigation (10 pp., listing 35 cases); and supervision and regulation of commercial banks (10 pp.). The rest of the *Report* provides income and condition statements for the Board and the twelve regional Banks; a capsule description of each merger, consolidation, acquisition of assets or assumption of liabilities approved by the Board in 1973; a *potpourri* of statistical tables related to *FR* operations, requirements, schedules, committees, and councils; and five tables providing data on the number and consolidated balance sheets of various classes of banks. There are, of course, no sections identifying legislative issues the Board prefers not to discuss or summarizing academic or business criticisms of *FR* policies.

After communicating the flavor of the Board's survey of last year's important economic events, our review focuses on the record of *FOMC* and *FR* Board domestic policy actions in 1973. For convenience, the macro-economic and micro-economic effects of these actions are discussed separately.

I. The Official Economic Story of 1973

Although the Fed's description of economic events is rich in empirical detail, even the System's best friends will find its account analytically superficial and overly politic. Just as the Council of Economic Advisers began its annual *Report*, the Board starts by recounting the severity of current inflation. This inflation is attributed to a host of exogenous forces that conscientious government agencies could not have offset without working even more serious harm on the economy. Conflicts between short-run and long-run policy benefits are not discussed. The impact of the February 1973 devaluation of the dollar is acknowledged, but the possibility that unwise past policies of any official body might have contributed to current inflation is not even mentioned. Nor is the possibility that fiscal policy (which "turned somewhat more restrictive in 1973") might have placed too heavy a stabilization burden on monetary policy. Wage-price controls, far from emerging as a Frankenstein monster mindlessly disrupting subjective probability distributions of future profits and wages, were wisely administered in 1973. However, the public's "attitude" toward the move to Phase III "contributed to the bulge in prices that ensued" (p. 5).

Rapid inflation naturally produced some micro-economic strains, notably in the housing and financial-institution sectors and in markets for bank loans, savings, and mortgages. To ease these problems, the Committee on Interest and Dividends (headed by *FR* Board Chairman Arthur Burns) introduced the dual prime rate and the Fed made timely use of moral suasion, adjustments in deposit-rate ceilings, and marginal reserve requirements on large CDs and selected nondeposit liabilities.

In the fall, the *FOMC* moved "to a moderately less restrictive monetary policy . . . warranted by the leveling-off in the economic expansion and by the evidence of developing weakness in the

economy caused by the oil shortage and other factors" (p. 10). Moving into 1974, "the job for monetary policy . . . is to steer a course that will not exacerbate present and prospective inflationary forces, but at the same time will avoid an unacceptably severe or extended weakening in economic activity that might develop from the energy crisis" (p. 11). Clearly a job for SuperFed!

II. The Fed and the Macroeconomy: *FOMC* Decisions in 1973

The burden of steering this difficult course falls primarily on the Federal Open Market Committee (*FOMC*). Monthly meetings of the *FOMC* establish the short-run thrust of monetary policy in exquisitely worded formulations of what is now called the "domestic policy directive." This directive is intended to guide the manager of the System Open Market Account (*SOMA*) in planning security transactions consistent with *FOMC* goals. The *SOMA* manager is further constrained by "continuing-authority directives" that limit the type and amounts of securities that he may trade.

Deciding whether and how to modify the delicate phrasing of last month's policy directive is the *FOMC*'s principal business. Before making this decision, the *FOMC* reviews a vast quantity of economic information: Commerce Department estimates of real output in prior quarters and staff projections of changes in output, prices, and sectoral expenditures in the current and succeeding quarter; monthly figures on retail sales, industrial production, employment, wages, and prices; statistics on foreign-exchange markets and *U.S.* trade and payments; and data detailing developments in financial markets. The Committee also hears a report on the nature and effects of last month's open-market operations. Against this background, the *FOMC* claims to take "note" of a staff analysis that cites numerical targets that the *SOMA* manager should aim at during the next month. As Table 1 indicates, these suggested targets (discussed more fully below) are reported only in part. Unfortunately, the record does not indicate what model and projections lead the staff to the specific range of policy settings it recommends. The staff must construct an outstanding case though, since (so far as we can tell) *FOMC* decisions depart from staff recommendations in relatively minor and carefully justified ways.

Evolution of the FOMC Directive, 1966 to 1974

Until 1966, the *FOMC*'s policy directives focused entirely on so-called money-market conditions.[1] It instructed the *SOMA* manager to supply or absorb reserves by open-market sales and purchases according to the state of some target subset of money-market variables. Although the particular subset varied over time, it typically included measures both of bank reserve positions and of short-term interest rates. In principle, open-market operations were used to establish or maintain optimal money-market conditions, which were supposed (through unspecified linkages) to achieve the nation's basic economic goals: desirable time paths for prices, employment, and output.

In practice, such money-market strategies left too much effective policy-making authority to the *SOMA* manager, who directed open-market operations primarily to smoothing the cyclical course of short-term interest rates. Linkages between short rates and goal variables proved sufficiently unreliable that critics came to call the strategy "money-market myopia." To remove the myopia, the money-market strategy has given way in four steps to

[1] Interesting discussions of the structure and formulation of these directives may be found in Jack Guttentag (1966, 1969), Stephen Axilrod, Andrew Brimmer, Sherman Maisel, and Ralph Young.

TABLE 1—1973 *FOMC* TARGETS FOR THE OPEN-MARKET ACCOUNT MANAGER

Date of Meeting	Stated Intentions with Respect to Growth in Monetary Aggregates	Consistent "Tolerance" or Target Range in Two-Month Per-Annum Growth Rate in *RPD* Suggested by Staff Analysis	*FOMC* Decision on "*RPD* Tolerance Range"	*FFR* Target Identified in Subsequent *FOMC* Policy Record
Jan. 16	"Slower . . . over the months ahead than occurred in the second half of last year."	9 to 11%	4.5 to 10.5%	$6\frac{3}{8}$%
Feb. 13	"Slower . . . over the months ahead than occurred on average in the past 6 months."	0.5 to 2.5%	−2.5 to 2.5%	7%+
Interim Telephone meeting (*ITM*) on March 1: Conferred about too-rapid growth in *RPD* and agreed to let the *FFR* rise to limit *RPD* growth.				
Mar. 19–20	Unchanged	14 to 16%	12 to 16%	7%
ITM on Apr. 11: Conferred about weaker-than-expected growth in *RPD* and aggregates, but declined to change policy guides so near the date of the next formal meeting.				
Apr. 17	"Moderate . . . over the months ahead."	"relatively rapid rate of growth"	10 to 12%	$7\frac{3}{4}$%
May 15	"Slower . . . over the months immediately ahead than occurred on average in the past 6 months."	"relatively rapid growth"	9 to 11%	8 to $8\frac{1}{2}$%
*ITM*s on May 24 and June 8: Conferred about too-rapid growth in *RPD* and aggregates, deciding to let the *FFR* rise if necessary to limit *RPD* growth.				
June 18–19	"Somewhat slower . . . over the months immediately ahead than appears indicated for the first half of the year."	9.5 to 11.5%	8 to 11.5%	$9\frac{3}{4}$%
ITM on July 6: Conferred about too-rapid growth in *RPD* and aggregates, deciding to let *FFR* rise if necessary.				
July 17	Same as June except for the deletion of "somewhat."	11.5 to 13.5%	11.5 to 13.5%	$10\frac{1}{2}$%
Aug. 21	"Slower . . . over the months immediately ahead than has occurred on average thus far this year."	13 to 15%	11 to 13%	$10\frac{3}{4}$%
Sept. 18	"Moderate . . . over the months ahead."	15 to 17%	15 to 18%	$10\frac{3}{4}$% to 10%
*ITM*s on Oct. 2 and 10: Conferred about slower than desired growth in *RPD* and aggregates and decided to pursue easier "money market conditions."				
Oct. 16	Unchanged	2 to 4%	2 to 5%	10%+
Nov. 19–20	Unchanged	−3 to −1%	−3 to −1%	$10\frac{1}{8}$%
ITM on Nov. 30: Conferred about too-rapid growth in aggregates, but decided, "to maintain current money market conditions for the time being."				
Dec. 17–18	"Some easing in . . . money market conditions, provided that the monetary aggregates do not appear to be growing excessively."	"moderate growth"	$8\frac{1}{4}$ to 11%	$9\frac{3}{4}$%
ITM on Jan. 11: Conferred about the possibility of *RPD* growth near upper tolerance limit and excessive growth in aggregates, but decided "to maintain the prevailing money market conditions for the time being."				

Note: RPD is Reserves Available to Support Private Nonbank Deposits and *FFR* is Federal Funds Rate.

what is now called a monetary-aggregates strategy. Observers suspect that these changes may involve a distinction without a difference.

The first step took place in June 1966, when the *FOMC* first added a "proviso clause" to its directive. This proviso clause alerted the *SOMA* manager that target money-market conditions would cease to be desirable if a single monetary aggregate —total bank credit (proxied by member-bank deposits)—developed outside of an agreed-upon range of values. But how he should act on this information was left indefinite.

Dissatisfied with the results, in January 1970 (at Chairman Martin's last *FOMC* meeting), the Committee decided that "increased stress should be placed on the objective of achieving modest growth in the monetary aggregates." Two months later, at the second *FOMC* meeting chaired by Arthur Burns, this decision was incorporated into the directive, instructing the *SOMA* manager to seek first and foremost a pattern of growth in a subset of monetary aggregates. In the short run, his job was to maintain "money-market conditions consistent with this objective." When the "intermediate" monetary variables (which monetarists believe to be linked reliably to goal variables) went off track, the *SOMA* manager typically set up an interim telephone meeting (*ITM*) with *FOMC* members to reassess the situation.

Because information on the growth of the aggregates was only available weekly (with nonmember-bank deposits measured only at call dates), neither of these developments much changed the information set by which the *SOMA* manager guided his day-to-day decisions. In hopes of establishing an effective day-to-day guide, in February 1972 the *FOMC* began to express its reserve objectives in terms of reserves available to support "private" deposits (*RPD*), defined as total member-bank reserves less reserves required on government and interbank deposits. Each month's *RPD* objective is formulated as a pair of control limits: A specific range of tolerable rates of growth in *RPD* over the month of and the month following the *FOMC* meeting. This *RPD* range is stated as part of the record of each meeting.

However, tolerance ranges intended for the federal funds rate (*FFR*), demand deposits adjusted plus currency outside banks (M_1), and M_1 plus net time deposits (M_2) were not reported publicly until the January 1974 meeting. Reporting this information is the fourth and latest step in the evolution of the directive. Although adopted primarily to conform more fully with requirements for public disclosure, this move may prove the most effective measure yet to tie the *SOMA* manager to an aggregates strategy.

Making these figures public should prove a boon for researchers and a source of embarrassment and on-the-job pressure for the *SOMA* manager. As argued in the next section, the *RPD* targets of 1972 and 1973 appear to have been set with the short-run needs of the money market very much in mind. The *RPD* targets have bounced around much more than the federal funds rate (*FFR*) and presumably more than the M_1 and M_2 targets. Publishing *RPD* targets has not embarrassed the Fed because *RPD* figures are neither widely published nor (given the present structure of reserve requirements) easy to interpret. Raw growth rates in *RPD* must be adjusted for "breaks" occasioned by changes in the level and structure of reserve requirements. Short of espionage, it is not possible to reproduce the precise adjustments made in the figures used by the *FOMC* prior to 1974. Publishing M_1 and M_2 targets and comparable figures on subsequent growth in these variables and *RPD* amounts to the *FOMC*'s stripping off its bureaucratic armor. Critics in Congress, in the financial press, and in aca-

deme should enjoy sorting out intended from unintended fluctuations in the money supply. If the *FOMC* can stand the heat, we should all learn a good deal.

FOMC Strategy in 1973: The More Things Change, The More They Remain The Same

In 1973, the Fed claimed to follow what it conceives as a three-pronged aggregates strategy.[2] The short-run focus is on the *FFR* and on the growth rate of *RPD*. Over longer periods, the short-run targets are supposed to be adjusted to insure satisfactory rates of growth in such monetary aggregates as M_1 and M_2. The idea is to make whatever open-market purchases and sales are necessary to keep the *FFR* and growth in *RPD* within targeted bounds, but to change these bounds (usually via an interim telephone meeting *ITM*) when and if monetary-aggregate growth rates move off target.

At best, this conception seems wishful. First, because of lagged reserve accounting, *RPD* growth cannot be controlled tightly over short periods of time. Except for minor variation in desired excess reserves and in interperiod averaging of reserve holdings, a floor is set under *RPD* growth by the growth in required reserves implied by the pattern of deposit change in the previous settlement period. In practice, the *FOMC* has not made its *RPD* target a binding constraint on the *SOMA* manager. For 1973, this is brought out in Tables 1 and 2. Table 1 summarizes 1973 changes in the domestic policy directive. Table 2 is designed to relate these to subsequent *RPD* growth and to emphasize that month-to-month variation in the target range for *RPD* is net of substantial seasonal swings. Growth rates in seasonally adjusted *RPD*, adjusted also for reserve-requirement changes made during the last seven months of 1973, should closely approximate the *FOMC* series.

As wide as the *RPD* tolerance range is, in eight of the twelve months, growth experience lies outside the designated range by more than $\frac{1}{2}$ point. This suggests that *FOMC* strategy is composed effectively of only two elements: tolerance ranges on the *FFR* and on monetary growth rates g_m. For future reference, we designate these target ranges as FFR^* and g_m^*.

Second, even this limited strategy seems unrealizable. The plan is to control a price in the short run and a quantity in the long run: to smooth interest-rate movements without sacrificing the attainment of other economic-stabilization goals. But for the *SOMA* manager, smoothing interest rates tends to be an *intermediate-run* objective (related loosely and uncritically to the System's panic-prevention responsibilities), whose attainment may compromise long-run goals. The desire to smooth interest rates accounts for much of the observed variability in the *RPD* tolerance range. The midpoint of this range moves like the transform of a series of desired changes in the *FFR* series.

Third, the two-element strategy is well-understood by professional investors.[3] Not only can banks and securities dealers observe changes in the location of the FFR^* range almost as soon as these occur, they can typically predict them in advance. Looking at published figures on money supply, deposit, and credit growth, they can usually ascertain whether the aggregates have been developing on or off the *FOMC*'s target paths. This knowledge justifies informed speculation that intensifies procyclical movements in the aggregates.[4] This speculation increases the dis-

[2] Other discussions of *FOMC* decisions in 1973 may be found in Sheila Tschinkel and Alan Holmes, and Anatol Balbach and Jerry Jordan.

[3] The argument that follows is adapted from William Poole.

[4] In 1974, speculation seems to have been triggered by changes in business loans at weekly reporting banks. In July 1974, the Board even took action to make sure that figures on money and on large CDs and business loans at New York City banks were released at the same time by each *FR* bank.

TABLE 2—DESIRED AND ACTUAL AVERAGE TWO-MONTH GROWTH RATES IN *RPD* DURING 1973, BY MONTH
(in percent per annum)

FOMC Meeting Month	*FOMC* Tolerance Range lower limit	*FOMC* Tolerance Range upper limit	Growth Rate in Total *RPD* Adjusted for Reserve-Requirement Breaks, *RPD**	Growth Rate in Seasonally Adjusted *RPD** (estimate of *FOMC* series)	Growth Rate in Unadjusted *RPD*
January	4.5	10.5	2.9	6.5	2.9
February	−2.5	2.5	−21.3	3.7	−21.3
March	12	16	10.3	10.1	10.3
April	10	12	11.4	9.9	11.4
May	9	11	5.9	13.6	5.9
June	8	11.5	16.9	17.9	23.4
July	11.5	13.5	14.3	14.2	28.6
August	11	13	11.0	11.6	18.8
September	15	18	10.7	7.1	18.4
October	2	5	2.0	−2.6	10.5
November	−3	−1	8.1	1.6	7.8
December	8.25	11	26.3	8.1	20.0
Mean	7.1	10.25	8.2	8.5	11.4
Standard Deviation	5.7	5.5	11.3	5.7	12.8

Note: All *RPD* series were compiled by Stephen H. Zeller of the Board's Staff. The *RPD** series is adjusted backward and multiplicatively for breaks in the *RPD* series caused by changes in the level and structure of reserve requirements. Unpublished *FOMC* adjustments for such breaks are additive and limited to the interval in which each reserve-requirement change occurs. In 1973 the seasonally adjusted *RPD** series tracks the pattern of qualitative deviations that led to *ITM*s except in December, a time when unadjusted *RPD* growth was high and reserve requirements were reduced.

crepancy between g_m^* and g_m. To see this, let us suppose that newly published figures on g_m and bank-credit expansion suggest that monetary expansion is faster than the Fed would want. Odds are good that the *FOMC* will soon raise the *FFR**. This expectation of a future rise in the one-day rate leads banks and dealers to switch their borrowing out of the one-day market and into longer markets (for example, by raising rates on large CDs) and also to reduce their borrowing by selling open-market assets (perhaps selling them short) before the interest-rate rise occurs. This raises rates in other short-term markets, putting upward pressure on the *current* *FFR* and extracting additional reserves from the Fed to defend the current *FFR**.

Over an entire boom, this pattern of speculation and temporary *FFR** defense makes the Fed an involuntary "engine of inflation" unless it forces subsequent monetary growth rates low enough to offset these preadjustment spurts in reserves. During 1973, monthly growth rates in M_1 and M_2 show only a small negative correlation (less than 0.3 in absolute value) with the observed deviation from the *RPD* tolerance range of the previous two months. This suggests that, much as under previous money-market strategies, an unintended procyclical element of reserve growth has remained part and parcel of the new strategy.

Mired perhaps in its old (interest-rate and credit) habits of thought, the *FOMC* viewed its stance as one of restraint over the first three-quarters of the year, with a switch to slightly easier conditions in the last three months to soften the impact of the energy shortage. Two-month growth rates in four seasonally adjusted monetary and reserve aggregates are compared in Table 3. Despite the *FOMC*'s professed concern for *RPD* and monetary aggregates, in 1973 the only intermediate-target variable that behaved consistently as if it were being controlled according to the *FOMC*'s

TABLE 3—COMPARISON OF AVERAGE TWO-MONTH GROWTH RATES IN ALTERNATIVE *FOMC* INTERMEDIATE-TARGET VARIABLES DURING 1973, BY MONTH
(in percent per annum)

FOMC Meeting Month	Simple Growth Rate in Seasonally Adjusted *RPD**	Compound Growth Rate in Seasonally Adjusted: Bank Loans and Investments	M_1	M_2
January	6.5	19.8	5.3	8.0
February	3.7	22.0	3.3	5.9
March	10.1	18.7	3.5	7.0
April	9.9	15.6	10.4	10.5
May	13.6	14.4	15.0	13.0
June	17.9	12.4	9.5	9.6
July	14.2	14.0	1.8	6.3
August	11.6	14.0	−2.0	5.2
September	7.1	9.2	0.7	7.6
October	−2.6	7.5	8.7	11.9
November	1.6	7.1	11.3	11.4
December	8.1	7.8	3.6	8.9
Mean	8.5	13.5	5.9	8.8
Standard Deviation	5.7	5.0	5.0	2.5

Source: The first column of figures is carried over from Table 2. The last three columns are taken from *Monetary Trends*, prepared by The Federal Reserve Bank of St. Louis. These figures reflect the May 1974 benchmark adjustments made by the staff of the *FR* Board.

announced intentions was the *FFR*. Whether monthly fluctuations in M_1 and M_2 average out as desired over six-month periods is debatable, but the pattern by quarters is decidedly stop-go. Even the growth in bank credit through August does not correspond with the operative words of the monthly *FOMC* directives.

That bank credit grew faster than M_1 and M_2 can be attributed to micro-economic incentives generated in large part by the regulatory framework the Fed has erected. As market interest rates rise, the costs of reserve requirements and ceilings on deposit rates rise, leading banks (and their customers) to substitute innovative new liabilities for traditional types of deposits. This process of substitution requires us to broaden the operative definition of what asset components constitute the money supply. For monetarists and nonmonetarists alike, M_1 and M_2 became less-adequate guides to the macro-economic impact of monetary policy.

III. The Fed and the Microeconomy: Effects on Relative Rates of Return

The *FR* manipulation of other monetary instruments (reserve requirements, discount rate and borrowing procedures, deposit-rate ceilings, and complementary jawboning) serves three overlapping purposes:

1. to underscore System open-market policy by generating supporting "announcement effects";
2. to alleviate micro-economic difficulties that current open-market operations may have aggravated;
3. to "perfect" the regulatory framework within which commercial banks operate.

In addition to their effects on the macro-economy, these instruments alter the vector of relative rates of return either across banks or across different categories of assets and liabilities. It is useful to view

these instruments of monetary policy as taxes.[5] Unlike open-market operations, adjustments in these variables seriously affect individual incentives and have qualitatively predictable impact effects on the distribution of income. Since these allocative and distributive effects are controversial, official *FR* statements describing manipulations of these instruments often take the bureaucratically safer course of ignoring their most important microeconomic effects and the Board's true motivation for undertaking an adjustment. Explicit balancing of distributive and allocative costs and benefits is not undertaken.

A. *Adjustments in Reserve Requirements*

Reserve requirements in excess of what a bank would voluntarily hold may be interpreted as a 100 percent tax levied on the income that would have been earned on funds put aside to meet the requirement. The existence of *FR* reserve requirements creates an incentive for commercial banks to withdraw or to refrain from joining the *FR* System. The incentive varies with the onerousness of reserve requirements applicable to comparable nonmember banks in the same home state and with the level of interest rates on earning assets. Increases in interest rates increase the cost advantage enjoyed by nonmember banks. Although it could, the Fed *does not* stabilize this incentive by setting a variable rate of interest on member-bank reserves. Differences in the level of reserve requirements on different types of liabilities (including zero requirements on innovative forms of debt instruments) lead member banks to favor the expansion of the less heavily taxed sources of funds. These few propositions explain the accelerating postwar erosion of System membership. They also explain the faster growth in bank credit than in M_1 and M_2 and a number of actions undertaken by the Board in 1973 as administrative responses aimed at reducing tax-avoidance opportunities.

1. *Actions to Restrict the Use of Exempt Debt Instruments.* Three offbeat types of bank debt not subject to reserve requirements grew to become serious sources of regulatory slippage: finance bills, standby letters of credit and other loan commitments, and so-called "other demand liabilities."

Finance bills are innovative instruments known also as "ineligible acceptances." They were exempted from ordinary regulations until July 1973 on the legal technicality that since the drafts being accepted did not arise from trade-related transactions, they were ineligible for discounting at *FR* Banks and therefore not mentioned in the *FR* Act. When reserve requirements were first imposed on them in July, about $1.5 billion of these marketable instruments were outstanding.

Standby letters of credit (also known as documented discount notes) are irrevocable contingent loans made to an issuer of commercial paper to guarantee repayment of the proceeds. As a contingent liability, the note does not appear on the balance sheet of the bank. Except for avoiding applicable reserve requirements and (until September 1974) limits on loans to a single borrower, the transaction is equivalent to a sale of commercial paper by the bank, with the proceeds loaned to the firm receiving the guarantee. The profitability of the transaction to the bank is rooted in the difference between the credit ratings of the customer and the bank. In April 1973, member banks were directed to maintain and make available to examiners records of the volume of loan commitments, estimated future takedowns, and the credit standing of the borrowers. Banks appearing to make unwise use of such commitments are to receive unspecified applications of moral suasion

[5] See Richard Posner.

to encourage them to revise their policy.

Finally, in November 1973, the Board redefined gross demand deposits. Although this action accords with longstanding efforts to speed up the check-clearing process and to improve money-supply statistics, its timing is explained by the Board's desire to stop some banks from lowering the effective level of required reserves. As interest rates rose, member banks found it increasingly attractive to place temporarily uncleared funds (from checks that had been debited against customer accounts but not yet settled with the presenting bank) into a so-called "other liability" account that had not previously been subject to reserve requirements. Within a local clearinghouse, as long as all banks delayed settlement and activity shares were similar, individual banks would not lose from this practice.

2. *Imposition of High Marginal Reserve Requirements of Hot-Money Liabilities.* Together with interest ceilings set under Regulation Q, the rising opportunity cost of reserve requirements led banks to raise progressively more funds from nontraditional sources. In turn, the volatility of some of these liabilities has increased the nation's potential for financial panic. A loss of confidence in commercial-paper issuers, such as occurred following Penn-Central's 1970 bankruptcy, would bring the banking system under enormous pressure today.

To provide countervailing incentives, the *FR* Board increased and refined reserve requirements on certain hot-money sources of funds several times during the year.[6] To limit the undesired allocational and distributional effects of these higher requirements, the Board took two unusual steps. First, it applied jawboning pressures to large nonmember and foreign-owned banks to enlist their "voluntary" cooperation in conforming to these requirements. Second, it formulated the hot-money reserve requirement as a *marginal* one. In this way, the tax would apply only to increments in relevant liabilities that occurred after a base date. This produced much the same stabilization effects with less impact on member-bank profits. In April 1973, a requirement of 5 percent was set on the base level, with a marginal requirement of 8 percent on increments. At the same time, the reserve requirement on foreign borrowings (primarily Euro-dollars) was reduced from 20 percent to that on other hot money. The marginal rate was raised to 11 percent in October, but dropped back to 8 percent in December after money-market rates stabilized at a lower level.

3. *The FR Board's New Plea for Universal Requirements.* The Fed's adoption of a graduated schedule of reserve requirements on demand deposits (in November 1972) was intended to reduce the incentive for small banks (for whom the benefits of System membership have traditionally been small) to quit the System. However, the continued rise in interest rates has more than offset this adjustment. As a result, when requirements on demand deposits were increased in July 1973 to underscore the *FOMC*'s anti-inflationary stance, the 8 percent requirement on a bank's first $2 million in deposits was left unchanged.

The Fed's continued interest in universal reserve requirements (uniform Federal Reserve requirements on all banks) traces both to its desire to eliminate the problem of eroding System membership and to its hunger for greater regulatory dominion. This interest is expressed again in the legislative-recommendations section of the Board's *Report.* However, the

[6] Hot-money liabilities now subject to reserve requirements include large-denomination CDs, bank-related issues of commercial paper, certain foreign borrowings of *U.S.* banks, and ineligible acceptances. Bank liabilities not yet subject to reserve requirements include standby letters of credit and forward transactions in foreign-exchange markets.

reasons given for proposing this legislation are disingenuously stated as "fostering equity among institutions" and "improving control over money and credit." What makes the argument disingenuous is the deference paid the dual banking system and the lack of any demonstration either that increasing implicit taxes on nonmember banks is better than reducing those on member banks or that this method of improving control is better than various alternatives (such as paying interest on reserves) that disinterested monetary theorists might propose.[7]

B. *Decisions on Discount-Rate Levels and Borrowing Procedures*

1. *Basic Policy.* Traditionally, official discussions of discounting and discount-rate decisions have buried the issues in humbug. While reserve requirements are a tax on member banks, the Fed's discount-rate policy has made borrowing an outright subsidy. One-day differentials between the discount rate and other money-market rates reached record heights in 1973. While the discount rate rose only as high as 8 percent, the prime rate reached 10, the rate on large CDs hit $12\frac{1}{4}$, prime commercial paper peaked at $10\frac{3}{4}$, and 91-day Treasury Bills rose to 9.01. The *FFR* was over 10 for most of five months, hitting a one-day peak of 26 percent.

These differentials made the subsidy element in *FR* borrowing increasingly attractive, forcing the *FR* Banks to ration discount requests energetically. *FR* discretion in granting borrowing requests became increasingly important, augmenting the leverage of System efforts at moral suasion and its ability to reduce the net costs of membership to complaining banks.

The justification the *Report* offers for allowing the discount-rate subsidy to reach these heights reinforces the popular misconception that the Fed "sets" interest rates and exemplifies the Board's traditional preoccupation with announcement effects—its inordinate fondness for sending out unmistakable *nonverbal* signals of its policy intentions. For example, after open-market rates peaked in the fall, the fear of precipitating perverse expectations effects led the Board to turn down regional *FR* Bank requests for higher discount rates, even though the level of rates remained substantially above the discount rate. To explain the Fed's unwillingness to tie the discount rate to market rates, one must look to the System's interest in offsetting rising implicit tax payments associated with reserve requirements. In 1973, a nasty person might wonder also about the temporary restraint on discount-rate adjustments introduced by the on-going cosmetic battle against prime-rate increases waged by Chairman Burns in his conflicting job as head of the Committee on Interest and Dividends.[8] Customer arbitrage of the difference between the prime and the rate on large CDs maddened the bankers. It would have been awkward for Burns to push up the discount rate, while demanding that banks hold back on the prime. But this restraint must have been a minor influence, since ceasing the battle against the prime has not produced freer movement in the discount rate.

2. *Seasonal Borrowing Privilege.* A seasonal borrowing privilege was established in April 1973, aimed at small banks (those "that lack reasonably reliable access to national money markets"). The stated purpose of these new arrangements is "to assist those banks in meeting seasonal needs for funds arising from a recurring pattern of movements in deposits and

[7] It is ironic that in July 1974 the U.S. Treasury (the ultimate beneficiary of the reserve-requirement tax on member banks) announced its plan to seek authority to shift part of the working balances it keeps at commercial banks into interest-bearing time deposits.

[8] For a detailed analysis of this battle and its effects on the size distribution of loans, see the author.

loans that persists for at least 8 weeks" (p. 77). But the Board's interest in dealing with this longstanding problem at this time surely lies in its desire to offset the growing net tax on small member banks, thereby reducing their incentive to quit the System. Each bank's seasonal line of credit is arranged in advance, with an initial maturity as long as 90 days. (I have heard of an effective maturity as long as 9 months.) This procedure allows the Fed to tailor its subsidy to the level necessary to convert an unhappy bank into a willing member. The basic fairness of such discrimination deserves to be discussed. It would seem economically rational, but much less efficient than paying interest on reserves.

C. *Panic-Prevention Taxes and Subsidies*

One of the Fed's primary obligations is to prevent temporary financial stringency from degenerating into capital-market crisis. Fed officials work toward this end in two ways: (1) macroeconomically, by smoothing short-run fluctuations in interest rates, and (2) microeconomically, by subsidizing weakened depository firms: (a) globally, through insulating them from direct competition for deposits (thereby taxing their customers and competitors), and (b) individually, through making credit available to troubled firms at bargain rates.

1. *Global Policy: The Fight To Keep Interest Ceilings Effective.* In the last nine years of accelerating inflation, micro-economic panic-prevention policy has been keyed to the plight of depository institutions, such as savings and loan associations (*S&L*), that place their funds predominantly in long-term instruments. With most of their assets carried over from a lower-interest past, these firms are particularly vulnerable to rapid run-ups in interest rates. Many of these firms could not generate enough current earnings to pay competitive rates of interest on savings deposits, while few possess more than a meager accumulation of capital on which to draw. Free competition for savings would force the weakest of these firms either to close their doors or to sell out to stronger competitors. Fed officials have feared that widespread bankruptcies or reorganizations would bring on full-scale panic in *U.S.* financial markets, an event which would pain *all* Americans.

To protect these institutions, the *FR* Board has cooperated in instituting and constantly mending a leak-prone system of comprehensive controls on deposit interest. This continuing policy has made it difficult for savers to earn a positive real rate of return on small accumulations of wealth. It has also been a source of continual headache to the Board. Because it is profitable for borrowers and lenders to seek ways around interest-rate ceilings, authorities can only bottle up market pressures over a prolonged period of time by extending the reach of their controls over an ever-widening portion of *U.S.* markets for loanable funds. To avoid this consequence, ceiling interest rates on savings deposits at banks and thrift institutions have had to be scrapped on large commercial-bank CDs (first in 1970 on short maturities and then on maturities over 90 days in May 1973). Ceilings on remaining savings instruments have had to be supplemented by restrictions on the minimum denominations of short-term government securities and by rumored strong-arm measures to prevent large retail organizations (Sears Roebuck and AT&T) from issuing deposit-like "mini-bonds" tailored to the needs of the small saver. Politically, it has been embarassing to have to explain continually that some "greater good" requires denying low-income and middle-income households access to the same high rates of return that

their rich and powerful neighbors enjoy.

It is far from clear either that the benefits of shackling small savers are as great as claimed or that these benefits could not be obtained at a lesser and more fairly distributed cost. With modern deposit insurance, the failure or reorganization of the *S&L* industry's weakest firms need not set off a falling-dominoes chain reaction of financial collapse. On the contrary, rationalizing the thrift industry should increase its overall ability to raise funds, thereby increasing the amount of mortgage money available.

When open-market interest rates are high, ceilings on deposit interest rates simultaneously intensify the disruption of mortgage lending and deprive small and unsophisticated savings depositors of interest income that is rightfully theirs. If thrift institutions must be subsidized, the costs should at least be calculated and placed before Congress for debate. Presenting figures of this sort in the Board's *Annual Report* could make it an important document.

a. *1973's Wild-Card Experiment*

Although consumer advocates have raised these issues repeatedly, the thrift industry (abetted by small banks) has persuaded Congress to pressure the Fed and other regulatory agencies into burying them again as quickly as possible. The process is illustrated by the political murder of the Fed-championed "wild-card experiment," which between July and October of 1973 allowed banks and thrifts to offer any interest rate they wished on certificates of deposit that met two conditions: (1) a minimum denomination of $1000 or more, and (2) a maturity of at least four years. Packaged with a $\frac{1}{2}$ to $\frac{3}{4}$ point increase in ceiling rates on other categories of time and savings deposits, these so-called topless 4's were "designed (1) to provide room within the ceilings for a greater measure of equity in the payment of interest to consumers, in an environment of generally rising interest rates, and (2) to enable member banks to bid more effectively for consumer deposits in competition with the yields available to savers on market securities" (p. 96). However, these beneficial adjustments were combined with an unfavorable tightening of the rules governing early withdrawal. The new rules required that depositors making an early withdrawal forfeit three months' interest and earn only the low passbook rate over the additional period (if any) for which their funds had remained committed. This effectively limited interest in the topless instruments to savers who could afford to put funds aside for a substantial amount of time.

A second unfavorable element applied initially only to *S&Ls*. Authored (perhaps as an act of sabotage) by the Federal Home Loan Bank Board (*FHLBB*), this consisted of a percentage limitation on the amount of these new instruments outstanding to 5 percent of an institution's total time and savings deposits. This limitation inhibited *S&Ls* from seeking topless funds energetically. But a number of banks made imaginative offers, the most successful at variable rates tied to future rates of inflation or yields on Treasury Bills. The combination of rising yields in the open market and spirited bank competition cost *S&Ls* over $2 billion in net outflows between July 1 and September 30. Commercial-bank competition via this instrument was slowed by *FR* Board rulings: (1) on July 26th, which imposed the 5 percent limitation on banks "to provide for the introduction of these new savings instruments at an orderly pace;" and (2) on August 23rd, which defined any adjustment in the interest rate of a time-deposit contract as an early withdrawal, thereby eliminating the variable-rate feature that had proven es-

pecially popular with small savers. It was slowed further by the conviction that *S&L* lobbyists had convinced Congress to vote the topless 4's back under a ceiling. This legislation passed in October, with congressional debate making wild-card CDs (and the agencies that favored them) the scapegoats for *all* the funding problems experienced by the thrift and housing industries during last summer's rapid run-up in open-market interest rates. As the Board's *Report* delicately suggests, even without the vehicle of the topless 4's, bank and open-market competition would have attracted a sizeable amount of funds away from thrifts in any case.

b. *Attempts to Circumvent or to Exploit Deposit-Rate Ceilings*

One can cite a number of developing institutions and practices which, with continuing *U.S.* inflation, threaten to skewer the very institutions that deposit-rate controls are meant to protect.

i. *Short-term money-market mutual funds, investment trusts, and custody accounts.* These institutions pool funds subscribed by small savers to let them overcome the barriers of high minimum denominations and odd-lot transactions costs. Some custody arrangements have even been set up by banks as a way of selling large CDs to customers who would otherwise place their funds elsewhere.

ii. *NOW accounts.* NOW stands for "negotiable order of withdrawal." In Massachusetts and New Hampshire, mutual savings banks located a loophole in the law that permitted them to issue what are effectively interest-bearing checking accounts. After a hard lobbying battle, in 1973 Congress voted to let all depository institutions in these two states offer such accounts "experimentally," while outlawing them elsewhere. Several New York savings banks have already found a loophole in the new law that allows them to offer checking accounts anyway, albeit at zero interest.

iii. *Repurchase agreements with household customers.* Since a repurchase agreement is not technically a deposit, it is exempt from reserve requirements and Regulation Q unless and until the *FR* Board rules to the contrary.

iv. *Floating-rate notes.* These instruments emerged in July 1974. As conceived originally by Citicorp, they had three features designed to appeal to the small saver: (1) a low minimum subscription of \$5,000 and \$1,000 denominations; (2) a variable interest rate tied to Treasury Bill rates; and (3) a semiannual option of redemption at par. Intense lobbying by *S&Ls* suggested the wisdom of modifying the third feature to make the new notes a bit less competitive with savings deposits. To get *FR* Chairman Burns to testify before Congress in favor of the new instrument, Citicorp agreed to lengthen the interval until the first semiannual redemption date to two years. Once the Citicorp issue was cleared, interest in offering these instruments spread quickly to nonbanking (and even nonfinancial) firms.

2. *FR Commitment to Bail Out Troubled Institutions Individually.* A deal for Fed support was possible primarily because of the growing threat of large-bank failures. As we have seen, large banks have financed their asset expansion less by growth in traditional types of deposits and more and more by borrowing in the short-term cap-

ital markets via innovative instruments whose implicit income-tax rates are (until the Fed extends its reserve-requirement and interest-rate regulations to capture them) relatively low. On balance, these policies have increased the volatility of large-bank liabilities. Nevertheless, in the last decade, large banks have allowed their average ratio of capital to risky assets to fall by about one-fifth. Many of these banks have also (for a healthy fee) guaranteed a substantial volume of commercial paper issued by firms whose credit ratings would not otherwise have qualified them to borrow in that market at favorable rates. If in a period of financial stringency these corporations prove unable to pay off this paper at maturity, these standby letters of credit would automatically convert these customers' debts into immediate claims on the assets of the guaranteeing bank.

Banks that have adopted these "modern" portfolio policies have increased their exposure to liquidity crises and necessarily become more dependent on the Fed to bail them out in the event of adverse market developments. Conversely, Fed officials must worry—more than they have had to worry since 1929—about the possibility of a chain of bank failures triggering a financial panic. Especially after the spectacular foreign-exchange losses and rapid CD and federal-funds runoffs experienced by the Franklin National Bank, the threat of bank failures weighed heavily on the minds of the Board and noticeably depressed the market for bank stocks.

Large banks' aggressive acceptance of greater interest-rate and default risks has effectively backed the Fed into a corner and underscored the potential conflict between the Fed's duty to fight inflation and its responsibility to maintain confidence in the liquidity and integrity of private financial institutions. In early 1974, the strength of the Fed's determination to slow the rate of inflation surprised many bankers, who speculated that in the first half of the year the Fed would ease interest rates to boost employment. Many large banks borrowed billions of dollars in one-day and other very-short funds to lend at typical maturities. When higher rather than lower interest rates developed, these institutions suffered substantial losses.

To maintain confidence in the financial system, Fed officials confirmed their intention to lend to troubled banks on an individual-needs basis. To prevent such loans from reducing the tightness of monetary restraint, Fed officials insisted that reserves injected in this way would be offset by sales from the Fed's securities portfolio. These promises put Fed policy on another contradictory path: a tightwire from which missteps are both easy and dangerous.

Moreover, the promise to help troubled banks on a needs basis raises treacherous problems of equity and incentives. How much should the Fed lend a troubled bank, at what rate, and for how long? When the federal-funds rate is averaging 12 percent, loans at an 8 percent discount rate represent an outright subsidy to the borrower. If anything, a troubled bank should pay an additional interest premium of at least 1 percent over the *FFR* to compensate for the special riskiness of its situation. Massive Fed lending actually transmits sure-fire earnings (or capital) to the borrowing bank.

Fed lending to Franklin National was rumored to be about $1.2 billion. (This is also roughly the amount of the surge in member-bank borrowing recorded in the week following the emergence of the crisis.) At our estimated interest differential of 5 percent, this loan represented a subsidy of $5 million a month.

How much help can the Fed provide Franklin National without making ordinary discount administration more dif-

ficult[9] and encouraging other banks to take even greater risks? If stockholders suspect that spectacular losses will—through Fed lending—be pushed back into the range of nearly average earnings, they should be more willing to see bank managements reach out for dangerous loans. In this respect, the Fed's proposed reconciliation of its competing responsibilities carries the seeds of its own destruction. While this course of action will provide short-term relief, in the long run it threatens to sharpen both spikes of the policy dilemma.

Given the existence of deposit insurance, what is needed is machinery that would allow a bankrupt bank to fail, while transferring its deposit liabilities (and perhaps even some of its labor force and physical capital) to other institutions. This would maintain the integrity of the payment system without rewarding bad management. It would also free the Fed to concentrate on macro-economic conditions.

IV. Summary Comment

The year 1973 was busy for the Fed. Although its problems were longstanding ones, they gained considerable urgency. Open-market policy is not noticeably slowing inflation, while inflation and deposit-rate controls are steadily distorting the financial structure. Commercial-bank avoidance of increasing implicit taxes associated with regulatory restraints poses many administrative problems and has increased banks' aggregate risk exposure. With two-digit inflation, market pressures to surmount deposit-rate controls threaten to keep *S&Ls* and mortgage markets under considerable pressure.

All the while, financial-reform legislation remains immobilized by the vigorous efforts of short-sighted lobbyists determined to preserve and even to augment the short-run legislative profits of the institutions they represent.

[9] Underscoring these problems, a *Boston Globe* cartoon pictured John Q Public eagerly bringing a depleted piggy bank to Chairman Burns.

REFERENCES

S. H. Axilrod, "The FOMC Directive As Structured in the Late 1960's: Theory and Appraisal," in Staff of FR Board of Governors, *Open Market Policies and Operating Procedures*, Washington 1971, 1–36.

A. Balbach and J. L. Jordan, "The Federal Open Market Committee in 1973," *Fed. Res. Bank St. Louis Rev.*, Apr. 1974, *56*, 2–17.

A. F. Brimmer, "The Political Economy of Money: Evolution and Impact of Monetarism in the Federal Reserve System," *Amer. Econ. Rev. Proc.*, May 1972, *62*, 344–52.

J. M. Guttentag, "The Strategy of Open Market Operations," *Quart. J. Econ.*, Feb. 1966, *80*, 1–38.

———, "Defensive and Dynamic Open Market Operations, Discounting, and the Federal Reserve System's Crisis-Prevention Responsibilities," *J. Finance Proc.*, May 1969, *24*, 249–63.

E. J. Kane, "Politicians Against the Prime: The Dual-Rate Fiasco," *Bankers Mag.*, Spring 1974, *157*, 88–96.

S. J. Maisel, *Managing The Dollar*, New York 1973.

W. Poole, "Benefits and Costs of Stable Monetary Growth," prepared for Univ. of Rochester Conference on Money, Unemployment, and Inflation, Apr. 1974.

R. A. Posner, "Taxation by Regulation," *Bell J. Econ.*, Spring 1971, *2*, 22–50.

S. Tschinkel and A. R. Holmes, "Open Market Operations in 1973," *Fed. Res. Bank New York Mon. Rev.*, May 1974, *56*, 103–16.

R. A. Young, *Instruments of Monetary Policy in the United States: The Role of the Federal Reserve System*, Washington 1973.

Federal Reserve Bank of St. Louis, *Monetary Trends*, June 20, 1974.

On the Irrelevance of Corporate Financial Policy

By JOSEPH E. STIGLITZ*

This paper extends to a multiperiod model, the argument of Franco Modigliani and Merton Miller and the author (1969) that the financial policy of the firm is of no consequence. In doing so, we are able to consider a much wider class of financial policies: not only does the firm choose a debt-equity ratio, but it also selects a dividend-retention ratio and a maturity structure of debt, and it may even decide on holdings of assets (securities) in other firms. I wish to show, in the context of a general equilibrium model, that none of these policies has an effect on the valuation of the firm, under certain seemingly weak circumstances. Whether these assumptions are "realistic" or not is a question of some debate, about which I shall have a few words to say later. But by clarifying the assumptions, I hope at least to focus the discussion on the relevant issues.

The question of the effect of firm financial policy on the valuation of the firm is obviously of central concern to students of corporate finance; if the conditions under which the "irrelevance" theorems obtain are deemed realistic, it robs them of much of their stock in trade. But the question of the irrelevance of financial policy is of far greater significance.

We can divide the decisions of the firm into four groups:

(a) How should the firm finance its investment?
(b) How should the firm distribute its revenue?
(c) How much should the firm invest?
(d) Which projects should the firm undertake (or what techniques of production should the firm employ)?

The first two decisions of the firm are the *financial* decisions of the firm, the latter two the *real* decisions. The theory of corporate finance focuses on the financial decisions. The two financial decisions are closely related (see below), and so are the two real decisions. What is not obvious is the relationship between the real decisions and the financial decisions. An answer to this question requires an analysis of the relationship between corporations and the household sector of the economy, and to further our understanding of this relationship is a primary object of this paper. If the hypothesis that the financial policy of the firm makes no difference to the firm's market valuation is correct, it also means that if firms maximize their market value, the *real* decisions are the only decisions that count, and the financial decisions have no bearing on them. In particular, it means that analyses of the real sector based on "flow of funds analysis"—and conclusions such as that of Nicholas Kaldor that because the flow of funds from the household sector to the corporate sec-

* Stanford University. This is a revised version of a lecture delivered at a conference at Hakone, June 25–26, 1970. I am very much indebted to Hirofumi Uzawa, who organized the conference, and to the other participants for the helpful comments and discussion. The research described here was supported by the Guggenheim Foundation, the National Science Foundation, and the Ford Foundation. Like all workers in the area of corporate finance, I owe a great intellectual debt to Merton Miller and Franco Modigliani; in addition, I have benefited greatly over the years from helpful discussions with them on the issues discussed here. I am also indebted to Robert Merton and Peter Kerbel for their helpful discussions.

tor is very small, the decisions of households with respect to savings are of relatively little significance in the determination of the equilibrium of the economy—are not likely to give us much insight into what is really going on: at best they provide us with some spurious correlations.[1] Moreover, if the maturity structure of debt is of no consequence, it casts some doubt about the validity of the partial equilibrium models attempting to relate the maturity structure to the term structure of interest rates (see, for instance, my 1970 paper).

In the literature, two different but closely related propositions have been confused: they both assert that the financial policy of the firm has no affect on its valuation. One asserts, however, that the individual is indifferent to alternative financial policies, in particular to debt-equity ratios, and hence there is no determinate debt-equity ratio for the economy as a whole. That is to say, any change in the financial policy of the firm can be completely offset by the actions of the stockholders (and indeed will be offset in the new general equilibrium situation).

The second proposition asserts that the individual may not be indifferent to alternative financial policies, that there may be for instance a determinate debt-equity ratio for the economy as a whole, but the financial policy of any *particular* firm makes no difference. The first asserts, in other words, the irrelevance of the financial structure for the entire economy, and *therefore* of the particular firm; the second only asserts the irrelevance of the financial structure of an individual firm. Clearly the former proposition is a much stronger one than the latter.[2] We are concerned here with both kinds of propositions.

The paper will proceed as follows. In Section I, the basic model is set up. In Section II, I prove my fundamental theorem on the irrelevance of the firm's financial policy from the point of view of any individual. Section III will comment briefly on the assumptions made and their limitations. Section IV will show that financial policy need not be of concern to any particular firm, even if it is of concern to individuals, under much weaker conditions than those used to demonstrate the earlier proposition.

I. The Basic Model

A. *Firms*

The various financial decisions of the firm are very closely related. One of the interests in a multiperiod model is to explore these relationships. A decision to increase the amount to be distributed as dividends means that if the firm were to leave its investment decision unchanged, it would have to raise additional revenue to pay for the planned investment. If it raises more by issuing bonds, the amount left over for distribution next period will be decreased, and hence either retained earnings or dividends next period must be reduced. If it raises the revenue by issuing shares, it means the amount distributed to each shareholder next period (if retained earnings were unchanged next period) would be reduced. Thus, the interrelations among all the decisions are complex and any decision today may have ramifications for many periods into the future.[3]

[1] E.g., the correlation between retained earnings and investment does not provide an explanation of the determination of the level of investment. See, for instance, John Meyer and Edwin Kuh.

[2] The Modigliani-Miller theorem was really of the latter type: they show that if there are two or more firms of the same risk class (the same pattern of returns across the states of nature), then the debt-equity ratio of any particular firm is indeterminate. The author's theorem (1969) was of the former type.

[3] The importance of these relationships has often been missed by even as astute students of the theory of corporate finance as William Baumol and Burton Malkiel, and Modigliani and Miller. In discussing the impact of taxation on the optimal financial policy of the firm, they observe that increasing the debt reduces the tax liability of the firm and hence increases its value.

For expositional simplicity, we shall use a "one-commodity" model;[4] each period there is a single commodity input and a single commodity output (dollars or yen). We shall look at the consequences of alternative financial plans on the firm's market valuation, given a "real plan" of the firm. A real plan is characterized by a statement of the investment level and choice of technique in each period contingent on the state of nature (the set of events that have occurred up to that time). Thus, given the real plan, we know the level of profits in each period, depending of course, on the state of nature. Let

$I_i(t, \theta(t), k)$ = the level of investment of the ith firm at time t, if the state of nature at that date is $\theta(t)$, under plan k.

$X_i(t, \theta(t), k)$ = the output or gross profits of the ith firm at time t, if the state of nature at that date is $\theta(t)$ under plan k.[5,6]

There are a number of alternative ways a firm may finance its investment:

(a) It can finance its investment with retained earnings or by issuing new securities.

(b) If it issues new securities, it can use a number of different financial instruments: common stock, bonds, preferred stock, convertible bonds, etc. Each of these financial instruments carries with it different contractual rights with respect to the distribution of the gross profits of the firm, and the part the owner of those instruments can play in the decision making of the firm. For instance, bonds yield a fixed sum in every state of nature except when the firm goes bankrupt, in which case the proceeds of the firm are divided among the bondholders. However, except when there is the distinct possibility of the firm not being able to meet its debt obligations, the bondholders generally have no voting rights in the management of the firm.

The return to a common stock, on the other hand, is variable—except when the firm goes bankrupt, in which case it is zero. A shareholder is entitled to receive a proportionate share of the dividends of the firm. The dividends, of course, depend not only on the real policy of the firm but on the particular financial policy chosen. To know the stream of returns, the shareholder must know both the real and the financial decisions of the firm. On the other hand, if our argument that financial policy is irrelevant is correct, then although changes in dividend policy affect the pattern of returns received by any single share of the firm, the individual is indifferent to these changes. The shareholder (like the bondholder) can sell his shares at any date and receive what he can for them. Finally, ownership of shares generally gives one a proportionate vote in the stockholders' meeting (although some firms also issue shares which do not have voting rights).

In the ensuing analysis, we shall assume for simplicity that there are only two classes of financial instruments, bonds and common shares.

For an ongoing firm with a given investment policy, increasing the debt-equity ratio implies that the firm retains less of its earnings and thus the capital gains will be smaller.

[4] The analysis for the multicommodity model is identical, except now a new set of financial decisions becomes available to the firm: it can denominate its bonds in terms of money or in terms of some other commodity, or some composite commodity (e.g., the cost of living index). Indeed, certain decisions which may appear to be "real" are in fact financial: when relative prices are uncertain, firms must decide on whether to buy futures (or hold inventories) of inputs or sell futures (or hold inventories) of outputs. In short, all such "hedging" decisions have (under the assumptions below) no affect on the market value of the firm.

[5] No loss of generality is had by interpreting I_i and X_i as vectors of inputs and outputs. In the proofs, we then need to replace I_i by $v \cdot I_i$ and X_i by $v \cdot X_i$, where v is the vector of relative prices (at time t, in state $\theta(t)$).

[6] Since in our simplification there are no other inputs, the output of the firm and its gross profit (before paying interest on debt, etc.) are identical. The modifications required when there are other inputs are straightforward.

(c) If it decides to issue bonds, it must decide on what maturity—one year, two year, etc.—and what the coupon rate will be. For simplicity we shall assume that bonds carry no coupons. Thus, a t period bond is a promise to pay in t periods 1 dollar. When it is issued, it obviously sells at a discount.[7] Let

$p(t, \tau, \theta(t))$ = the price at time t in state $\theta(t)$ of a bond which promises to pay 1 dollar at time τ.

If there is uncertainty, the individual will not know what the price of such a bond will be in future periods except that, if there is no bankruptcy,

$$p(t, t, \theta) = 1 \qquad \text{for all } \theta, t$$

In the discussion below all variables are state-contingent, but for notational simplicity we omit the $\theta(t)$ except when it would otherwise be confusing. Similarly, all real variables (X_i, I_i) are dependent on the "plan k," but k too will be suppressed.

The relationships among the various financial decisions are expressed by the two accounting identities: Total investment must be equal to the value of the change in outstanding bonds plus the value of the change in outstanding shares plus retained earnings:

$$(1)\quad I_i(t) = \sum_{\tau=t+1}^{\infty} [p(t,\tau)(B_i(t,\tau) - B_i(t-1,\tau)) + q_i(t)(S_i(t) - S_i(t-1))] + RE_i(t)$$

where

$B_i(t, \tau)$ = the number of bonds outstanding at the end of period t with maturity at time τ.

$q_i(t)$ = the price of a share of the ith firm at time t.

$S_i(t)$ = the number of shares outstanding at the end of the period.

$RE_i(t)$ = retained earnings.

There is, of course, no natural unit for shares, so it is just as simple to define

$$E_i^+(t) = q_i(t)S_i(t)$$

as the value of the shares outstanding at the end of the period, while

$$E_i^-(t) = q_i(t)S_i(t-1)$$

as the value of the shares outstanding at the beginning of the period, i.e., the value at t-th period prices of the shares outstanding at the end of the previous period.

Thus $E_i^+(t) - E_i^-(t)$ is the value of the change in the number of shares outstanding resulting from issuing new shares during the t-th period; this should not be confused with $E_i^-(t+1) - E_i^+(t)$, which is the change in the value of those shares outstanding at the end of the t-th period from the t-th to the t+1st period. The latter is the capital gain (or loss) on existing shares.

The second accounting identity states that total income in state θ at time t must be equal to the income distributed (to bondholders and to shareholders) plus that retained by the firm.[8]

$$(2)\quad X_i(t) = B_i(t-1, t) + D_i(t) + RE_i(t)$$

where $D_i(t)$ are the dividend payments to stockholders on record at the beginning of the period; i.e., each share receives $D(t)/S(t-1)$ or the dividend per dollar invested at time t−1 is $D(t)/E_i^+(t-1)$.

Figure 1 illustrates a flow of funds diagram for this economy over time. (Because the flow of funds occurs over time, the diagram is not made circular.) It should be noted that we have drawn the line for retained earnings through the

[7] There are of course still other financial decisions, including ownership claims on other firms, what numeraire to denominate bonds in, etc. All of these could be included in our model—at some expense in notational complexity.

[8] Recall that we are assuming for notational simplicity that there are no coupons on bonds; thus bondholders only receive income from the firm upon maturation of bonds.

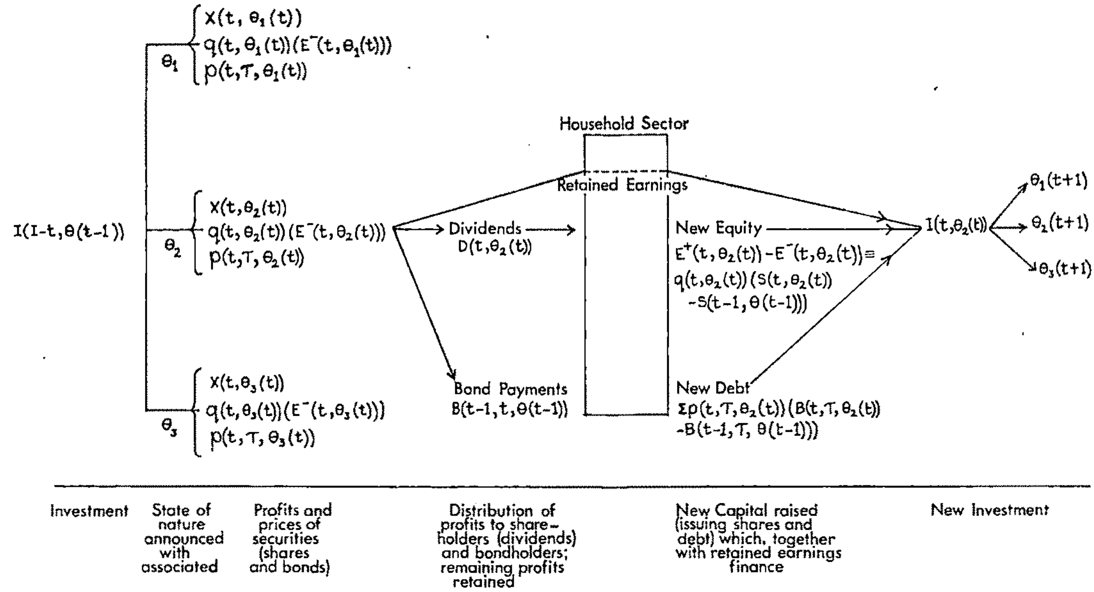


FIGURE 1

household sector with a dotted line: the fact that the earnings do not pass through their hands directly does not necessarily mean that the household sector does not include (in some sense) such retained earnings in their income. As the analysis below shows, the retained earnings will be directly reflected in the price of outstanding shares.

The diagram also serves to clarify the timing implicit in our analysis: Let us break into the diagram at a point where the firm has just made its "new" investment decision having raised the requisite capital. The output (profits) next period (which depends not only on investment in the period just ended, but on investment in all preceding periods, as well as the specification of the environmental path for these preceding periods) is unknown; we await the specification of the "environment" for time t, for example, the rainfall, temperature, etc. The "state of nature" is then announced, i.e., $\theta(t)$ is then given. This means that the set of possible outcomes for t+1 and beyond has immediately been substantially narrowed. In Figure 1 we can, for instance, now completely ignore all but one of the environmental paths passing through t.

Given the new information embodied in the announcement of the state $\theta(t)$, the value of the shares and the prices of bonds are determined. In particular, the value of the equity of the firm now is $E_i^-(t, \theta_2(t))$. Moreover, at this point, for the particular plan we have denoted by k, we know exactly what the firm plans to do this period: we know its investment, $I(t, \theta_2, k)$, how much dividends it plans to give out, how much it plans to retain, how many bonds and of what maturity it plans to issue, how many new shares it plans to issue, etc. We still don't know, of course, what its investment will be in the future; for this we await further information. But we do assume that we know what the firm will do in each contingency.[9]

[9] The assumption that the raising of new capital follows (in each period) the distribution of the profits and the bond payments is made simply for expositional convenience. In fact, these two operations may be (over)

In the theorem presented in the next section, two concepts play a crucial role: one is the value of the firm, the other is bankruptcy.

The total value of the ith firm is the *present* value of its outstanding bonds plus the value of its equity: at the beginning of the period (before maturing bonds are redeemed), this is

$$\text{(3a)}\quad V_i^-(\mathrm{t}) = E_i^-(\mathrm{t}) + \sum_{\tau=\mathrm{t}}^{\infty} p(\mathrm{t}, \tau) B_i(\mathrm{t}-1, \tau)$$

while the value of the firm at the end of the period is

$$\text{(3b)}\quad V_i^+(\mathrm{t}) = E_i^+(\mathrm{t}) + \sum_{\tau=\mathrm{t}+1}^{\infty} p(\mathrm{t}, \tau) B_i(\mathrm{t}, \tau)$$

Using (1) to (3), we can solve for the value of dividends in terms of the change in the value of the firm, its gross profits, and its investment:

$$\text{(4)}\quad D_i(\mathrm{t}) = X_i(\mathrm{t}) - I_i(\mathrm{t}) + V_i^+(\mathrm{t}) - V_i^-(\mathrm{t})$$

Bankruptcy is somewhat more difficult to define. The basic notion is, of course, that the firm is unable to meet its debt obligations. In the two-period model discussed in my 1969 paper, a firm is bankrupt whenever the profits are less than the nominal claims of bondholders.

$$X_i < (1 + \hat{r}_i) B_i$$

where $\hat{r}_i$ is the nominal rate of interest on the bond. If r^* is the nominal rate of interest on a perfectly safe bond, and $\min X_i(\theta)$ is the minimum profit in any state of nature, the probability of bankruptcy is zero provided

$$B_i \leq \frac{\min X_i(\theta)}{1 + r^*}$$

while it is positive if

$$B_i > \frac{\min X_i(\theta)}{1 + r^*}$$

See Figure 2. The analogous statement here would be

$$\text{(5)}\quad X_i(I_i, \theta(\mathrm{t})) < B_i(\mathrm{t} - 1, \mathrm{t}, \theta(\mathrm{t} - 1))$$

But this will not do. For firms always have the option, if their returns in a particular period in a particular state are low, of borrowing more or issuing new shares to meet these debt obligations. Indeed, this is exactly what they would normally do, provided their future prospects of returns are sufficiently good. Therefore the condition stated in (4) is at best a statement about very short-term liquidity, not about the solvency of the given firm. In the last period of a multiperiod model with a terminal date, the condition for bankruptcy is given by (5), since the firm cannot (by assumption) issue new shares or borrow further. But there is no reason to restrict ourselves to a finite period model.

Clearly, what we mean by bankruptcy is that at some date, in some state of nature, the value of the maturing bonds of a firm is less than the face value

$$p(\mathrm{t}, \mathrm{t}, \theta(\mathrm{t})) < 1$$

for some t and $\theta(\mathrm{t})$. This is equivalent to saying that at that date and in that state of nature the value of the equity of the firm is zero (or negative if there is not limited liability).[10]

$$\text{(6)}\quad E_i^-(\mathrm{t}, \theta(\mathrm{t})) \leq 0 \qquad \text{or}$$

$$\text{(6}'\text{)}\quad V_i^-(\mathrm{t}, \theta(\mathrm{t})) \leq \sum_{\tau} B_i(\mathrm{t} - 1, \tau, \theta(\mathrm{t} - 1)) \cdot p(\mathrm{t}, \tau, \theta(\mathrm{t}))$$

thought of as occurring simultaneously. The important assumption is that the same price of a security prevails at the beginning of the period as at the end. Since once the state of nature is announced, everything that will occur during that period is known, this is not an unreasonable assumption; alternatively, if we think of the distribution of returns and the raising of new capital as occurring simultaneously, this is clearly the appropriate assumption.

[10] Clearly, if the price of a share is zero, the firm cannot issue more equity to pay off the debt holders.

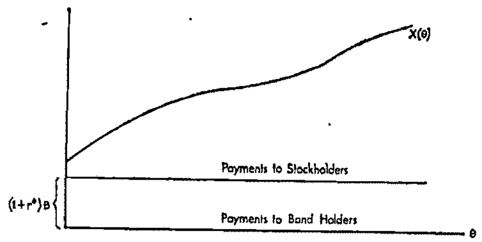


FIGURE 2a. DISTRIBUTION OF RETURNS: NO BANKRUPTCY

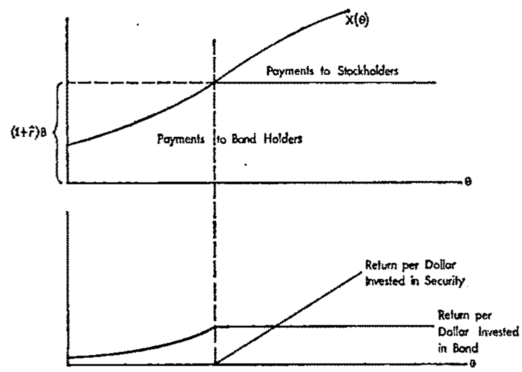


FIGURE 2b: DISTRIBUTION OF RETURNS: BANKRUPTCY

The fact that the value of the equity of the firm in some state of nature at some date in the future is zero does not mean that the value of the equity will be zero today; if there is some chance that the firm will not go bankrupt, clearly the value will be positive. But it does mean that bonds issued with maturities at the date of potential bankruptcy or beyond are risky securities, i.e., their terminal value is uncertain, and clearly the price of these bonds will not be the same as the price of a bond whose terminal value is certain. A change in the financial policy which results in a chance of bankruptcy, i.e., in a chance that the firm will not be able to meet its debt obligations, thus changes the prices of the bonds the firm issues, and it is the invariance of the price of the bonds which will be crucial in the argument of the next section.

We shall assume in the subsequent analysis that there is no bankruptcy.

B. *Households*

Let $w^{j-}(t)$ denote the jth individual's wealth at the beginning of the period and $w^{j+}(t)$ be his wealth at the end of the period. If $E_i^{j+}(t)$ is the individual's ownership of equity in the ith firm at the end of the period and $B^j(t, \tau)$ his ownership at the end of the period of bonds maturing at date τ (because of the no bankruptcy assumption, the bonds issued by all firms are identical), then

$$w^{j+}(t) = \sum_i E_i^{j+}(t) + \sum_{\tau=t+1}^{\infty} p(t, \tau)B^j(t, \tau)$$

Hence, using (3), we obtain

$$(7) \qquad w^{j+}(t) = \sum_i \alpha_i^j(t)V_i^+(t) + \sum_{\tau=t+1}^{\infty} p(t, \tau)\left(B^j(t, \tau) - \sum_i \alpha_i^j(t)B_i(t, \tau)\right)$$

where $\alpha_i^j(t) = E_i^{j+}(t)/E_i^+(t)$ is the fraction of the equity of the ith firm owned by the jth individual at the end of the t-th period.

At the beginning of the next period his portfolio is worth

$$(8) \quad w^{j-}(t+1) = \sum_i \alpha_i^j(t)V_i^-(t+1) + \sum_{\tau=t+1}^{\infty} p(t+1, \tau)\left(B^j(t, \tau) - \sum_i \alpha_i^j(t)B_i(t, \tau)\right)$$

Finally, since during the period he will receive dividends equal to $\sum_i \alpha_i^j D_i$, if he consumes $c^j(t)$, his end of period wealth will be related to his beginning of period wealth by the equation

$$(9) \qquad w^{j+}(t) = w^{j-}(t) - c^j(t) + \sum_i \alpha_i^j(t-1)D_i(t)$$

Thus, substituting (4), (7), and (8) into (9), we obtain

$$(10)\quad c^j(\mathrm{t}) = \sum_i \alpha_i^j(\mathrm{t}-1)\{X_i(\mathrm{t}) - I_i(\mathrm{t}) + V_i^+(\mathrm{t}) - \sum_{\tau=\mathrm{t}}^{\infty} p(\mathrm{t},\tau)B_i(\mathrm{t}-1,\tau)\} + \sum_{\tau=\mathrm{t}}^{\infty} p(\mathrm{t},\tau)B^j(\mathrm{t}-1,\tau) - w^{j+}(\mathrm{t}), \quad \text{all } j, \tau$$

Equations (7) and (10) can be thought of as defining the individual's consumption opportunity set. Given any set of ownership of shares $\alpha_i^j(\mathrm{t}-1)$ and bonds $B^j(\mathrm{t}-1,\tau)$ at t−1, (10) defines the value of consumption plus end of period wealth at t. Given the value of wealth at the end of period t, (7) defines the possible ownership of shares and bonds at t, which in turn defines the value of consumption plus end of period wealth for t+1. Note that the opportunity locus does not depend on either dividends or retained earnings; the intuitive reason for this will be made clearer in the discussion following the theorem in the next section.

C. *General Equilibrium*

Market equilibrium requires the total value of ownership claims on the ith firm equal the value of its equity, i.e.,

$$\sum_j E_i^{j+}(\mathrm{t}) = \sum_j \alpha_i^j(\mathrm{t})E_i^+(\mathrm{t}) = E_i^+(\mathrm{t}) \quad \text{or}$$

$$(11)\quad \sum_j \alpha_i^j(\mathrm{t}) = 1 \quad \text{for all t}, i$$

Similarly, demand for bonds of each maturity must equal the supply:

$$(12)\quad \sum_i B_i(\mathrm{t},\tau) = \sum_j B^j(\mathrm{t},\tau) \quad \text{for all t}, \tau$$

II. The Basic Theorem

We are now prepared to state and prove our central theorem.

THEOREM 1: (a) *Assume there is no bankruptcy of any firm in any state of nature.* (b) *Assume that there is a perfect market for perfectly safe bonds of all maturities. (By perfectly safe, we mean that the amount that they pay upon maturity is known for certain; the price of these different maturities at all other dates may be highly variable.)*[11] (c) *All firms have already made their real decisions (i.e., the value of k for each i is given).* (d) *Assume there is a general equilibrium, with all markets clearing, characterized by a given price in each state of nature at each time for each maturity bond and each firm having a given valuation in each state of nature and at each time t, and a given financial policy (i.e., a specification of debt-equity ratio, retention ratio, maturity structure of bonds).*

Then, there is another general equilibrium solution where any firm (or group of firms) has changed any (or all) of its financial policies, but in which the value of the firm and the price of all maturities of all bonds are unchanged (for all periods and states of nature), and investors have made offsetting portfolio adjustments, i.e.,

$$(13)\quad \Delta B^j(\mathrm{t},\tau) = \sum_i \alpha_i^j(\mathrm{t})\Delta B_i(\mathrm{t},\tau) \quad \textit{all } \mathrm{t}, \tau, j$$

i.e., each investor alters his holdings of bonds by exactly his stockholder's share of the change in debt of each maturity of all firms and

$$(14)\quad \Delta\alpha_i^j(\mathrm{t}) = 0 \quad \textit{for all } i, j, \mathrm{t}$$

or equivalently

$$(14')\quad \Delta E_i^{j+}(\mathrm{t}) = \alpha_i^j(\mathrm{t})\Delta E_i^+(\mathrm{t})$$

Each investor changes his equity holdings

[11] From a *consumption* point of view, it cannot be said which maturity is the "safer" bond, which is the riskier; i.e., it cannot be said the long-term bonds are riskier than short-term bonds. See my 1970 paper. What we shall show here is that from the point of view of the *firm*, the maturity structure is irrelevant.

of each firm in proportion to the firm's change in total equity capital.

The argument of the proof is simple. I shall show that if the value of the firm and the price of all maturities of all bonds are unchanged (for all periods and states of nature) then the set of consumption possibilities available to any individual is unchanged. Since the set of consumption possibilities is unchanged, the individual will choose the same consumption path (i.e., the same plan of consumption over time, which is clearly a contingent plan, depending on which states of nature occur). To do this, he changes his investment-portfolio allocation plan as described by (13) and (14). Finally, I show that if the set of investment-portfolio allocation plans originally adopted by the different individuals in the economy (that is, before the firm changed its financial policy) was an equilibrium, so that each date and in each state of nature markets cleared, then the set of new investment-portfolio allocation plans also constitutes an equilibrium.

PROOF

1. The consumption opportunity set is unchanged. Consider any feasible consumption path and its associated portfolio allocation. From (7), it immediately follows that if $w^{j+}(t)$, $p(t, \tau)$, and $V_i^+(t)$ are unchanged after a given change in financial policy of a firm, then the changes in portfolio allocation described by (13) and (14) are feasible; and from (10) if these changes are undertaken, then $c^j(t+1)+w^{j+}(t+1)$ is unchanged. Thus, if $c^j(t+1)$ is unchanged, $w^{j+}(t+1)$ is unchanged. Clearly, if the values of the firm and prices of bonds are unchanged, the value of w^j for dates before the contemplated change in financial policy begins is unchanged.

These statements immediately imply that a consumption stream that was feasible in the original situation is feasible with the changed financial policy of the firm, and conversely.

2. Since the consumption opportunity set is unchanged, if a particular consumption path $\{c^{j*}(t, \theta(t))\}$ is preferred to all other consumption paths in the original situation, then it is preferred in the new situation. Thus, the indicated changes in portfolios are not only feasible, but optimal.

3. If in the original situation, all markets cleared at each point of time (for each state of nature), they also do in the new situation. Since $\alpha_i^j(t)$ are unchanged, clearly if $\sum \alpha_i^j(t)=1$ before, it still does, and all markets for securities clear (equations (11) above are all satisfied). The change in demands for bonds of a given maturity at a given date is given by

$$\sum_j \sum_i \alpha_i^j(t)\Delta B_i(t, \tau) = \sum_i \Delta B_i(t, \tau)$$

i.e., it just equals the change in supply, so that if demand equaled supply in the original situation (equation (12) was satisfied) it still does.

III. Comments on the Theorem and Its Proof

There are four kinds of comments which I have to make. In Section IIIA we provide an intuitive interpretation of the theorem. In Section IIIB we point out how much weaker the assumptions employed in our analysis are than those used in previous proofs. In Section IIIC we discuss briefly the limitations on the proof, and how critical they are for the general validity of the theorem. In Section IIID we discuss the competitive forces at work to eliminate the "inefficiency" resulting from the resource allocation to financial management.

A. *Intuitive Interpretation*

The basic argument of the theorem is that individuals can exactly "undo" any financial policy undertaken by the firm.

Let us consider verbally what actions of the individuals are required to offset various actions by the firm. Assume the firm decreases its dividend payout ratio. This means that it has more retained earnings, so, if the two basic financial accounting identities are to be satisfied, either it must borrow less (perhaps it even lends) or issue fewer new shares. To make up for the loss in dividends, i.e., to keep the same consumption path, individuals buy fewer new shares in the firm or buy fewer new bonds. Assume the firm simply issued fewer shares. In one case, the value of the equity grew because of issuing new shares, in the other case, the value of the equity grew because of retained earnings. From the point of view of the stockholders, the two are perfectly equivalent. This change in dividend pay-out ratio thus leaves the debt-equity ratio unchanged. On the other hand, if the firm decreases the number of bonds issued, it will lead to a lower debt-equity ratio. Then individuals borrow on their own account. One can think of it as if the individual takes the proceeds of the loan to purchase the increased equity in the firm (since the two are exactly equal, this is only a convenient way of looking at it; since all funds are fungible, there is no real connection between the two). The increased borrowing by individuals exactly offsets the decreased borrowing by firms so markets continue to clear. Similarly, if the firm decides to issue more three-year bonds and fewer five-year bonds, the individual can undertake exactly offsetting actions in his own portfolio.

B. *On the Generality of the Theorem*

(a) *Risk classes, Arrow-Debreu securities, mean-variance analysis.* It should be emphasized that in the proof, it is not assumed that there are two or more firms which are otherwise identical; the argument does not require the existence of risk classes, as many states of nature as securities, or the assumption that returns can be described by means and variances, assumptions which have been crucial in other proofs of the more limited theorem on the irrelevance of debt-equity ratios.

(b) *Competitiveness of capital market.* No assumption about the competitiveness of the capital market has been made; the only assumption is that there be no discriminatory pricing, i.e., the price paid by one individual (firm) for a bond (or share) be the same as for all other individuals. But the market rate of interest—and hence the interest rate paid by a firm—may be affected by the amount of capital it raises from the market.

(c) *Rationality of consumers.* The only restriction on individual behavior is that given a set of feasible consumption paths, he always selects the same consumption path. Thus, the individual may maximize his discounted expected utility, but no such restrictive assumption is required for the result to obtain.

(d) *"Control" of firm.* Even if the individual does care about his political power (control) within the firm (which he may if the real decisions of the firm depend on the stockholders), if the role of each stockholder in decision making is simply a function of the proportion of the total shares he owns, the financial policy makes no difference, since political power of any shareholder is identical in the two situations.

One might have argued that a smaller equity base would make a "take-over" more likely; but under the assumption of no bankruptcy, this would not be true, since the group taking over the firm could borrow on the strength of the equity in the firm as collateral; if in the low equity situation, the group taking over could raise the requisite capital for a take-over,

they would have no problem doing so in the high equity situation.[12]

(e) *Source of uncertainty.* No assumption about the source of uncertainty is required.[13]

(f) *Multiplicity of equilibrium.* Theorem 1 is a theorem about market equilibria. It states that there are an infinite number of general equilibrium solutions of the economy all of which are identical in all respects except for the financial policies of firms and the value of bonds and shares (separately) held by individuals (although the proportions of the shares of each firm owned by any individual are the same). There may of course be more than one general equilibrium solution to the economy at any given set of financial policies. As usual, very strong conditions would be required to ensure uniqueness. But what our theorem does assure us is that if there are two (or three or . . .) equilibria at a given set of debt-equity ratios, then there are two (or three or . . .) at any other set of debt-equity ratios. The theorem has nothing to say about the important question as to which one of these will in fact be chosen.

(g) *Differing expectations.* The argument of the proof does not require that individuals have the same expectations. The only agreement in expectations that is required is that the firm will not go bankrupt in any state of nature. (See below for a discussion of this assumption.)

(h) *Market clearing.* The particular path described in the above analysis is an equilibrium path where individuals make plans all of which are *consistent* with one another, i.e., they are market clearing. In fact, the only thing required for the analysis is market clearing at time 0. In making his portfolio-consumption decision for time 0, the individual must have expectations of prices and firm valuations at all future dates and states. These may not, of course, be realized; at each successive date expectations are then revised. It is important to the analysis that these revisions depend on "real events" not on the financial structure of the firm (see below).

C. *Limitations on the Theorem*

There are three critical limitations on the theorem.

(a) *Independence of expectations from financial policy.* Our analysis requires that these *expectations* be unchanged as the firm changes its announced financial policy for the future.

If it should turn out that these expectations are a function of the financial policy of the firm, then in fact the financial policy of the firm will affect its valuation this period. The expectations that financial policy will affect market valuation are, at least in this very rough sense, fulfilled. But note that the argument for equilibrium paths shows that there is no reason that these expectations ought to change.

(b) *Individual borrowing an imperfect substitute for firm borrowing.* Perhaps the major objection to the proposition that the firm's financial policy is irrelevant is that individual borrowing is not a perfect substitute for firm borrowing. There are four principle reasons for this: 1) higher interest rates for individual borrowing than for

[12] For a more extended discussion of the relationship between debt-equity ratios, bankruptcy, and take-overs, see my 1972 paper.

[13] In particular the distinction between technological uncertainty and price uncertainty, which played such an important role in Diamond's analysis, is of no consequence here. It should also be noted that Diamond's assertion that his results do not depend on the no-bankruptcy assumption is incorrect.

corporate borrowing; 2) limitations on the amount that individuals can borrow from the market; 3) transactions costs; and 4) special tax provisions (differential treatment of capital gains and deductibility of interest payments for the corporate income tax). I have discussed these limitations in greater detail elsewhere (1969, 1973). Here I wish only to make a few observations.

First, the higher nominal interest rates individuals pay and the quantitative restrictions on their borrowings are primarily a reflection of the higher probability of default on the part of individuals. They are, in other words, a particular manifestation of the general problems that the chance of bankruptcy brings to the analysis.

Second, the first three reasons given above place restrictions on the set of financial policies among which the individual is indifferent, but there is no reason to believe that these restrictions are very severe. They may mean that firms cannot have all-equity policies, but individuals will still be indifferent among a wide set of debt-equity ratios. If, for instance, a firm were to decrease its debt-equity ratio, the analysis does not require that individuals borrow from the market to purchase the additional shares issued by the firm; it only requires that they decrease their holdings of bonds. Hence so long as the total debt of those firms whose shares the individual owns is sufficiently large that the individual is a net lender rather than a net borrower, the individual is indifferent. This places a lower bound on the "average debt-equity ratio" of the firms in the individual's portfolio (although not on the debt-equity ratio of any single firm). This constraint may become an important constraint if at those debt-equity ratios there is a finite probability of bankruptcy; that is, it is only in conjunction with the bankruptcy constraint that this constraint becomes significant.

Third, it does not place restrictions on the debt-equity policy to be pursued by any particular firm, only the set of debt-equity policies that groups of firms can follow; i.e., even if the constraint is binding, in general one firm can increase its debt-equity ratio when another firm decreases its debt-equity ratio. One cannot speak of an optimal debt-equity or optimal retention ratio.

(c) *Bankruptcy*. In my judgment, the most restrictive assumption is that of no bankruptcy.

The careful reader may have wondered where the restriction of no bankruptcy was used in the proof. Because of limited liability laws, it is clear (as I noted before) that $E_i \geq 0$. If the firm issues a sufficiently large number of bonds so that in some state of nature at some date $\sum pB_i > V_i$ for V_i to be the same as in the original (reference) situation, E_i would have to be negative. But this is impossible.

The assumption was not only critical to the proof, but I would argue, critical to the general validity of the theorem. To put it one way, it is not reasonable to assume that the price of bonds for which there is a positive probability of default at maturation would be the same as a perfectly safe bond. One might argue that the decline in the nominal value of bonds is compensated for by an equivalent increase in the value of equity, and under certain circumstances—the existence of as many securities as states of nature or the mean-variance model with homogeneous expectations—this is true. But in the more general case, bankruptcy changes the opportunity set facing a given individual so that the value of the firm is changed. Not only is the financial policy of importance, but no separation between the financial and real decisions is possible.[14] (See my 1972 paper.)

[14] These remarks should serve to clarify the difference between my theorem, both its meaning and its proof,

D. *Competitive Forces to Eliminate "Waste" of Resources on "Financial Management"*

One might ask, if financial policy is really of no importance, why do firms waste resources on "money managers"—shouldn't competitive forces lead all firms to ignore financial policy? Since worrying about it costs resources and can't increase the market valuation of the firm, clearly firms that spend resources on financial management have lower profits to distribute to their stockholders than firms who don't. There are five answers to this:

1) I have ignored some important considerations, in particular, taxes, which do make it profitable to worry about financial structures. Does this mean that I believe that in the absence of taxation financial managements would wither away? Not necessarily, or only very slowly, as the remaining points argue.

2) I have already argued that if individuals believe that financial policy affects firm valuation, then it will, and the firm that ignores the popular "prejudices" may do worse than one which takes them into account. There may have been no rational reason for the prices of tulip bulbs to rise as they did in the tulip bulb mania, but since they were rising, at least in the short run, one could make a "profit" by investing in them (see below).

3) Moreover, this relationship between the firm's financial policy and expectations about profits may not be as "irrational" as the above analysis suggests. Changes in financial policy may be an important signal for the "real prospects" of the firm. This would not be the case in my model, because in it there is no such thing as a liquidity crisis; but in the real world bankruptcy may be important, and the fact that banks and other lending institutions are unwilling to lend the firm money (for instance forcing a reduction in dividends to meet the liquidity requirements) may be a signal that those who know more about the prospects of the firm than the relatively uninformed shareholder are not sanguine about the prospects of the firm.[15]

4) There is, moreover, no reason that in the short run the different valuations lead to any inconsistencies or more generally that there are any forces leading individuals to reformulate their expectations so that valuations are independent of financial policies. Even if we have two firms which are identical in every *real* respect (that is, they belong to the same risk class, in the terminology of Modigliani and Miller), there is not necessarily any method by which individuals can arbitrage (over any short- or medium-run[16] period).[17]

and that of Modigliani and Miller. They assert that the financial policy is of no consequence. But Modigliani and Miller made use of risk classes in their proof, the use of which seemed to imply objective rather than subjective probability distributions over the possible outcomes. The mechanism which ensured that the debt-equity ratio made no difference for the value of the firm was individual arbitraging among different firms in the same risk class. Such arbitrage does not play any role in my analysis. Moreover, their argument was based on partial equilibrium analysis rather than general equilibrium analysis. It was not clear from their analysis whether the theorem held only for competitive markets, nor how the possibility of firm bankruptcy affected their results. The basic insight of the M-M analysis, that individual leverage could substitute for firm borrowing, remains the basis of my argument.

[15] Indeed, one might argue that this signalling effect of financial policy is one of its more important functions. If firms never issued dividends, simply retaining earnings (even in the form of bond purchases) then it might be possible for firms to postpone letting shareholders know when they are in "bad straits" even longer than they do at the present. This may provide part of the explanation of why, in spite of strong tax advantages for not issuing dividends, firms continue to do it.

[16] This qualification is imposed because, under certain circumstances, it can be shown that if different financial policies are pursued with the firms having different valuations and equal returns to the individual, then, in *finite* time, the *relative* valuations must become infinite. But finite, in this context, may be very long indeed. Such differences in valuations are (at least mathematically) very similar to the speculative booms (or depressions) which often seem to characterize price movements on the stock market. For a general discussion of these problems in a slightly different context, see Karl Shell and the author.

[17] To see this in the extreme case, we need only con-

5) Finally, we note that the resources wasted on financial management may be relatively minor (relative, say, to total profits of the firm) and hence the "competitive forces" to eliminate this inefficiency may operate with relatively little strength.

IV. Irrelevance of Financial Policy of Any Particular Firm

The above proposition established the irrelevance of the financial structure of the economy as a whole. The crucial assumption employed was that of no bankruptcy. We can remove this assumption and prove a weaker theorem about the irrelevance of the financial policy of any particular firm.

THEOREM 2: *Assume there is a general equilibrium for the economy which is characterized by a given market rate of interest (on safe bonds), by a given nominal rate of interest on the risky bonds of each of the firms which faces a chance of bankruptcy, and by each firm having a given market valuation and a financial policy (dividend-retention ratio, maturity structure of debt, etc.), and in which a given fraction of the shares of the firm are owned by the i-th individual.*

Let any firm (or any group of firms) change its financial policy. If financial intermediaries may be established costlessly, then there exists a new general equilibrium solution for the economy with the same market rate of interest, in which every firm has exactly the same market valuation as before, and in which the proportion of each firm's shares owned by the i-th individual, either directly or indirectly through intermediaries, is exactly the same as before.

Since the argument for changes in debt-equity ratio is perfectly analogous to changes in other financial policies, I shall focus my remarks on the debt-equity ratio. Assume in the initial equilibrium, there were no financial intermediary purchasing bonds and shares of the given firm. The firm changes its debt-equity ratio. A financial intermediary is created which reconstitutes the firm, i.e., purchases all of its bonds and shares, then issues bonds and shares in exactly the same ratio as in the original situation. The opportunity set facing the individual is completely unchanged, and hence the market valuations, rates of interest, etc. are completely unchanged.[18]

sider the situation where profits minus investment are known for certain and the firm issues no new shares. Then dividends for, say, firm 1 may be written (in a continuous time formulation)

$$D_1(t) = X_1(t) - I_1(t) - rB_1(t) + \dot{B}_1$$

where all bonds are assumed to be short-term bonds, earning an instantaneous rate of return of $r(t)$. Assume there are two firms which are identical except that one issues no bonds at all, i.e.,

$$D_1(t) = X_1(t) - I_1(t)$$

The total rate of return from owning shares in either company is simply the sum of dividends and capital gains, which we require to be the same for both. Dropping the subscripts on X and I, we have

$$\rho = \frac{\dot{E}_1}{E_1} + \frac{D_1}{E_1} = \frac{\dot{V}_1}{E_1} + \frac{X - I}{E_1} - \frac{rB_1}{E_1} = \frac{\dot{V}_2}{V_2} + \frac{X - I}{V_2}$$

or

$$\frac{\dot{V}_1}{V_1} - \frac{\dot{V}_2}{V_2} = (r - \rho)\frac{B_1}{V_1} - \frac{(X - I)}{V_1} + \frac{(X - I)}{V_2}$$

It is clear that if $\rho = r$, and $V_1 = V_2$ initially, then the two firms have the same value forever. But note that the second equation can also be satisfied with $\rho = r$ and $V_1 \neq V_2$, in which case there will be cumulative changes in the ratio of the valuations. Eventually, these will probably lead to one of the firms having an unusually low or unusually high earnings-valuation ratio, and this will probably lead to a revaluation of the firm. But there is no reason (without perfect futures market), as we have argued in detail elsewhere, that this might not go on for a long time.

[18] More generally, assume the original debt-equity ratio of the firm is d, and the kth financial intermediary purchases α^k of the bonds and equity of the firm and issues a debt-equity ratio of d^k. Thus $1 - \sum \alpha^k$ is the proportion of the firm purchased directly by individuals (not through intermediaries). Now assume the firm changes its debt-equity ratio to d'. Then all intermediaries except the one for which $d^k = d'$ are unaffected. It now issues a debt-equity ratio of d, and purchases $1 - \sum \alpha^k$ of the bonds and equity of the firm.

One might argue, however, that the opportunity set has been changed, because in principle the individual can buy bonds and shares in the firm directly as well as through the intermediary. Thus, if there is a probability of bankruptcy, his opportunity set is larger now than it was before. This may lead to an increase in the demand for the securities (bonds and stocks) of the given firm, so that the new situation is not an equilibrium one. There will be a new general equilibrium situation, with the value of bonds and equities of the firm greater than before. But this would imply that the original situation could not have been an equilibrium. For a financial intermediary could have purchased and issued the same fraction of the bonds and stocks of the given firm, thus obtaining a given fraction of the income of the firm in every state of nature, and then issued bonds and shares in the ratio of the debt-equity ratio of the "new" situation. The organizers of the intermediary would have then made a pure profit for themselves, equal to the difference between the value of the firm in the two situations.

The point of the argument is the following: If, corresponding to a given set of real and financial decisions of the other firms, the financial decision of the firm does make a difference, free entry of financial intermediaries will ensure that a set of financial securities will be marketed which maximizes the value of the firm regardless of the debt-equity ratio of the firm.

If one took the assumption of costless creation of intermediaries seriously, there is no reason to suppose that the process of proliferation of intermediaries would stop short of creating as many securities as states of nature; in which case, not only is the financial structure of any individual firm of no consequence, the financial structure of the economy is irrelevant.[19]

[19] That is, all financial structures that provide as many securities as states of nature are equivalent.

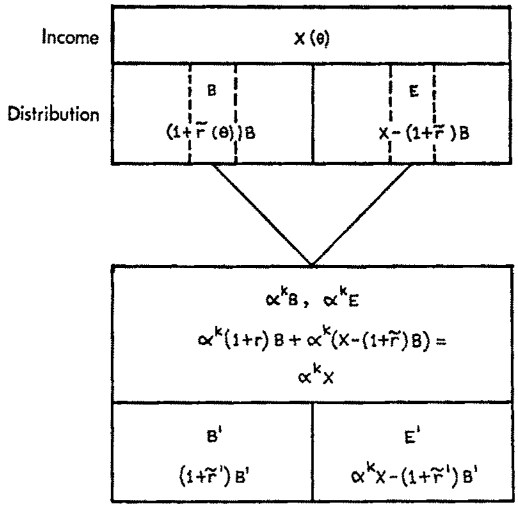


FIGURE 3. OPERATION OF FINANCIAL INTERMEDIARY ($\tilde{r}(\theta)$ = interest payments on bonds in state θ)

The fact that we do not see such a proliferation of financial intermediaries suggests either that they are unnecessary (that is, Pareto optimality can be obtained with a limited number of such intermediaries,[20] or the conditions of Theorem 1 are satisfied so that the financial structure is of no consequence), or that there are significant transactions costs relative to the gains to be had by the creation of such intermediaries.[21]

V. Concluding Comments

We have established the irrelevance of financial policy under a fairly general set of conditions. Three classes of limitations were noted to our result:[22] 1) expectations of real returns dependent on firm financial

[20] As in those circumstances in which the portfolio separation theorem is valid. See David Cass and the author.

[21] Included in "transactions costs" are the cost of obtaining information about different securities. If there is a finite probability of bankruptcy, purchasers of bonds have to evaluate the riskiness of the bonds. Thus not only does bankruptcy result in transactions costs when it occurs, but the potentiality of bankruptcy results in transactions costs at the time the bonds are sold.

[22] Besides the obvious distortionary effects of taxation.

policy; 2) individual borrowing not a perfect substitute for firm borrowing; 3) bankruptcy. Whether these limitations are important in practical applications is a moot question. But whether they are or are not, the theoretical importance of the theorem is not diminished: an understanding of it is to corporate finance as an understanding of the frictionless surface is to the understanding of the physics of motion. For some practical problems, friction can be ignored in a first approximation; in others it cannot. But even when it cannot, an understanding of what would happen in the absence of friction is essential. The empirical testing of the model is another matter: in physics we can attempt to approximate a frictionless surface and observe motions under those conditions; to do the analogous thing here would require us to create a world without transactions costs, tax distortions, and other frictions, and see if in these circumstances firms ignored their financial structure. It is, of course, essentially impossible to do this. Fortunately, the issue is not whether under those circumstances the financial policy would be irrelevant—most of us would agree that it would be—the issue is: how significant are the "limitations" and in what way do they affect corporate financial policy. The tests performed so far—such as examining the value (per unit scale) of firms thought to be essentially identical except for their debt-equity ratio—do not discriminate between worlds in which Theorem 1 is valid, those in which Theorem 2 is valid but not Theorem 1, or worlds in which neither theorem is valid (in which financial policy is important, but in which value-maximizing firms have selected the set of financial policies which maximize the firm's valuation). What is required is a greater understanding of the implications of these limitations and more refined tests to discriminate among the alternative hypotheses.

REFERENCES

W. Baumol and B. G. Malkiel, "The Firm's Optimal Debt-Equity Combination and the Cost of Capital," *Quart. J. Econ.*, Nov. 1967, *81*, 547–78.

D. Cass and J. E. Stiglitz, "The Structure of Investor Preferences and Asset Returns, and Separability in Portfolio Allocation: A Contribution to the Pure Theory of Mutual Funds," *J. Econ. Theory*, June 1970, *2*, 122–60.

P. Diamond, "The Role of a Stock Market in a General Equilibrium Model with Technological Uncertainty," *Amer. Econ. Rev.*, Sept. 1967, *57*, 759–76.

N. Kaldor, "Marginal Productivity and the Marco-Economic Theories of Distribution: Comment on Samuelson and Modigliani," *Rev. Econ. Stud.*, Oct. 1966, *33*, 309–19.

J. R. Meyer and E. Kuh, *The Investment Decision: An Empirical Study*, Cambridge 1957.

F. Modigliani and M. H. Miller, "The Cost of Capital, Corporation Finance, and the Theory of Investment," *Amer. Econ. Rev.*, June 1958, *48*, 261–97.

K. Shell and J. E. Stiglitz, "The Allocation of Investment in a Dynamic Economy," *Quart. J. Econ.*, Nov. 1967, *81*, 592–609.

J. E. Stiglitz, "A Re-Examination of the Modigliani-Miller Theorem," *Amer. Econ. Rev.*, Dec. 1969, *59*, 784–93.

———, "A Consumption Oriented Theory of the Demand for Financial Assets and the Term Structure of Interest Rates," *Rev. Econ. Stud.*, July 1970, *37*, 321–51.

———, "On Some Aspects of the Pure Theory of Corporate Finance, Bankruptcies and Takeovers," *Bell J. Econ.*, Autumn 1972, *3*, 458–82.

———, "Taxation, Corporate Financial Policy, and the Cost of Capital," *J. Publ. Econ.*, Jan. 1973, *2*, 1–34.

Unemployment, Inflation, and Monetarism

By JEROME L. STEIN*

Edmund Phelps criticized the game plan of the Council of Economic Advisers on the grounds that they had no explicit model of the *U.S.* economy to justify the policies that were undertaken to curb the inflation.

> The Council's Report provides a look at the 1969–71 game plan to disinflate by means of retarding aggregate demand. . . . The President-Elect's Task Force on Inflation recommended, as a first interim step, that aggregate demand be slowed so as to bring the unemployment rate back to some equilibrium region around 4.5 percent. . . . What happened was that, under the cover of the expectation and acceptance of such a limited step towards re-equilibration, the Administration gradually tightened monetary and fiscal policy so severely as 'gradually' to send the unemployment rate whizzing past the equilibrium zone to around 6 percent. To my knowledge the theory of how, and how well, this medicine would act to cure the patient of his inflation was never spelled out by the Council of Economic Advisers. [pp. 533–34]

Since the Council had no explicit model of the *U.S.* economy, they could neither explain the paradox of unemployment cum inflation nor could they evaluate the probable effects of alternate economic policies.

The econometrics of wage and price determination have been actively studied in recent years. Important articles on this subject appear in the *Brookings Papers on Economic Activity* and in the *Proceedings* of the Federal Reserve-Social Science Research Council (*SSRC*) Conference (edited by Otto Eckstein, papers by Hymans, Klein, de Menil and Enzler, Hirsch, Bodkin, Ball and Duffy, Eckstein and Wyss, Heien and Popkin, and Andersen and Carlson, among others). The *Proceedings* were analyzed by James Tobin in his summary comments. These studies generally contain three equations: a wage change equation, a price change equation, and an equation describing the formation of price expectations. The unemployment rate is taken as an exogenous variable; and the models do not explicitly contain monetary and fiscal policy variables.[1] Therefore, they cannot explain (i) the simultaneous determination of, and interactions between, the unemployment and inflation rates and (ii) the effects of monetary and fiscal policies upon the paths of the unemployment and inflation rates. The Andersen and Carlson study is an exception, however; it does contain the unemployment rate as an endogenous variable as well as the monetary and fiscal policy variables. It can and has been used to simulate the effects of policy. However, that model cannot be solved analytically; and its theoretical foundations are neither clear nor widely accepted.[2]

* Brown University. I am indebted to W. Bomberger, G. H. Borts, M. Friedman, J. Frenkel, H. Hori, E. F. Infante, N. Liviatan, and J. P. LaSalle for their perceptive criticisms of an earlier draft. The John Simon Guggenheim Foundation provided financial assistance for this study, which was written while I was at the Hebrew University in Jerusalem.

[1] Tobin summarized the *FRB-SSRC* models in this way: "These four equations form a subsystem of a complete model that given an initial price history is capable of determining prices, price expectations, and wages. Unemployment, past and present, and rental cost of capital can be taken as exogenous to the subsystem" (1972a, p. 6).

[2] Two studies of the relation between unemployment

Paul Samuelson wrote that "The central issue that is debated these days in connection with macro-economics is the doctrine of monetarism" (p. 7). However, there is no satisfactory theory of monetarism. Milton Friedman's attempt to provide a theoretical explanation of the monetarist position has been deemed unsuccessful both by neo-Keynesians and by other leading monetarists. Tobin described Friedman's theoretical framework as follows:

> He undoubtedly hoped that the use of a common theoretical apparatus would reduce the controversy about the roles of monetary and fiscal policies to an econometric debate about empirical magnitudes. If the monetarists and the neo-Keynesians could agree as to which values of which parameters in which behavior relations imply which policy conclusions, then they could concentrate on the evidence regarding the values of those parameters. I wish that these articles had brought us closer to this goal, but I am afraid they have not. [1972c, pp. 852–53]

According to Karl Brunner, who is a leading exponent of monetarism:

> Monetarist ideas have become almost respectable in recent years and the central propositions about the role of money and monetary policy have been increasingly accepted or at least seriously pondered. Still, the monetarist position remains an unsettled issue with respect to both its degree of confirmation and its analytic formulation. [p. 2]

Brunner and Allan Meltzer agree with Tobin's criticism of Friedman's article on monetarism: "We regard Friedman's discussion as either misleading or a complete reversal of his often stated position" (p. 846).

My paper attempts to explain the phenomena of unemployment cum inflation and the analytic foundations of the monetarist position within the context of a single model.

In Section I, I develop a simple small-scale dynamic model of the economy, whose salient features are accepted by most macro-economic theorists. The state variables are denoted by X, and it is the vector containing the unemployment rate, the rate of price change, and the expected rate of price change. Control variables denoted by vector C are the monetary and fiscal policies followed by the government. The dynamic model described by

$$DX = AX + BC, \qquad D \equiv d/dt$$

is easily solved analytically because it only contains three linear differential equations with constant coefficients. Every equation can be understood very easily, and the dynamics of the system and effects of policies can be clearly traced. Its steady-state and dynamic properties are discussed in Section II. Section III explains the phenomenon of unemployment cum inflation, and Section IV explains the logical foundations of monetarism, on the basis of the model contained in Sections I and II. A simple phase diagram is used as a graphic device to explain the qualitative implications of the dynamical system. Quantitative accuracy can only be obtained by estimating the coefficients of the dynamic model. Preliminary work along this line is encouraging, but is not reported here. My exclusive concentration in this paper is upon the theory.

I. A Simple Dynamic Model of the Macroeconomy

In the short run, the ratio of effective labor supplied per unit of capital is assumed to be relatively constant; and the state variables are the unemployment rate (U), the rate of price change (π), and the expected rate of price change (π^*). This section develops the complete model applicable to an economy where the ratio

and inflation must also be mentioned: David Laidler and the paper by the author and Ettore Infante.

of effective labor supplied per unit of capital is also a state variable. In the subsequent parts, only the short-run version will be used.

A. *The Labor Market and the Unemployment Equation*

The Walrasian labor market adjustment equation (1) states that the rate at which nominal wage W changes is a linear combination of two elements: the expected rate of price change π^*, and the state of the labor market.

$$(1)\quad \frac{DW}{W} = \pi^* + \lambda_1\left(\frac{N_d - N_s}{N_s}\right); \quad \lambda_1' > 0$$

No dichotomy is made between expectations of employers and of employees in this single sector macro-economic model; hence π^* reflects both the expected change in the price of output and the expected change in the price of goods purchased by consumers.

The state of the labor market is reflected by the second term $\lambda_1(\cdot)$. Variables N_d and N_s refer to the quantities of labor services demanded and supplied, respectively. Function λ_1' is positive; the growth of the nominal wage DW/W is positively related to the excess demand for labor $(N_d - N_s)/N_s$ for the usual reasons.

If the state of the labor market is given, variations in the expected rate of price change by employers and employees lead immediately to corresponding changes in the growth of the nominal wage.

Assume that technical change is Harrod-neutral, so that "effective labor" is a multiple $A(t)$ of labor in natural units. Then equation (1) can be written as (2) where x_d and x_s refer to effective labor per unit of capital demanded and supplied, respectively. Ratio $N_d/N_s = x_d/x_s$ when technical change is labor-augmenting.

$$(2)\quad \frac{DW}{W} = \pi^* + \lambda_1\left(\frac{x_d - x_s}{x_s}\right)$$

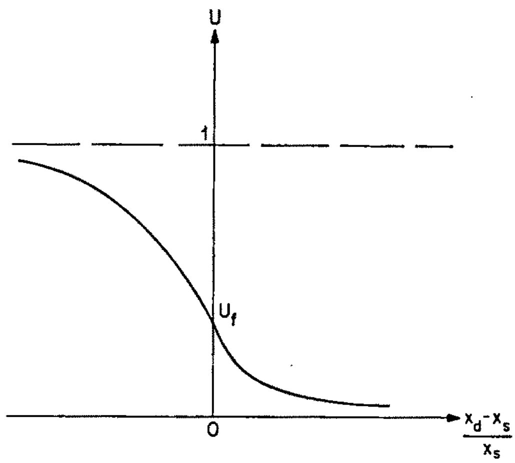


FIGURE 1. THE UNEMPLOYMENT RATE IS NEGATIVELY RELATED TO THE EXCESS DEMAND FOR LABOR

The excess demand for labor is not directly observable, though the closest direct measure of it may be the number of vacancies less the number of people unemployed. Equation (3) states that the measured unemployment rate U is negatively related to the excess demand for labor; and it is described in Figure 1. This assumption permits both positive and negative excess demands for labor to be associated with a positive unemployment rate. It does not require that we assume that the quantity of labor employed is the smaller of the quantity demanded or supplied. At full employment (see Figure 1), the unemployment $U_f \cdot N_s$ is equal to vacancies. Equation (3) is sufficiently general to be consistent with a wide range of micro-economic theories of labor markets. The inverse function is described by equation (4).

$$(3)\quad U = H\left(\frac{x_d - x_s}{x_s}\right), \quad H' < 0; \quad 1 \geqq U \geqq 0$$

$$(4)\quad H^{-1}(U) = \frac{x_d - x_s}{x_s}, \quad (H^{-1})' < 0$$

The actual ratio of effective labor em-

ployed per unit of capital, denoted by x (without a subscript), is equation (5) below.[3]

$$x \equiv (1 - U)x_s \tag{5}$$

If equation (4) is substituted into (2), then the frequently used wage change equation (6) is derived. Function h is the composition of λ_1 and H^{-1}, i.e., $h \equiv \lambda_1 \circ H^{-1}$, and h' is negative. Equation (7) is a linearization of (6).

$$\frac{DW}{W} = \pi^* + h(U), \qquad h' < 0 \tag{6}$$

$$\frac{DW}{W} = \pi^* + h_0 - h_1 U \tag{7}$$

The excess demand for labor per unit of capital (equation (8) below) requires some explanation. The demand for labor arises from the private and the government sectors. In the private sector, the quantity of labor demanded per unit of capital depends upon the real wage w adjusted for the level of technology $A(\mathrm{t})$. Specifically, the quantity of labor demanded by the private sector per unit of capital depends upon $w/A(\mathrm{t})$. Some of the output will be purchased by the private sector and some will be purchased by the government. The government also purchases the services of labor directly: the real gross product originating in the government sector is the real value of the compensation of employees. It is identically equal to a fixed multiple of the labor input of government employees, because gross product in constant prices per man-hour in the government sector is defined (by the Department of Commerce) to be unchanging over time. The government is not a profit maximizer in the conventional sense. Its direct demand for labor services per unit of capital G_1 can be taken as a control variable differing in magnitude from its purchases of goods from the private sector. Assume for simplicity that G_1 is a fraction ξ of total real government purchases of goods and services per unit of capital G; i.e., $G_1 = \xi G$. In general, ξ will not be a constant, and both ξ and G will constitute control variables. Therefore, the quantity of labor demanded per unit of capital[4] depends upon $w/A(\mathrm{t})$ and ξG, and is $x_d = F(w/A(\mathrm{t}), G)$.

The supply of effective labor per unit of capital x_s is assumed to be inelastic with respect to the real wage, and in the "short run," ratio x_s is assumed to be relatively constant. Therefore, the *excess* demand for labor per unit of capital, resulting from the activities of the private sector, is related to $w/A(\mathrm{t})$ the adjusted real wage. Combining the private and government sectors equation (8) is derived.

$$\frac{x_d - x_s}{x_s} = f\left(\frac{w}{A(\mathrm{t})}, G; x_s\right) \tag{8}$$

$$f_1 < 0, \quad f_2 > 0$$

Substitute equation (4) into (8) and derive equation (9). This shows the dependence of the unemployment rate U on the adjusted real wage $w/A(\mathrm{t})$ and on the real value of government purchases of goods

[3] The implication of (3) and (5) is that the ratio of employment to the quantity demanded $N/N_d = x/x_d$ is

$$\frac{x}{x_d} = \frac{1 - H\left(\frac{x_d}{x_s} - 1\right)}{x_d/x_s}$$

At any instant of time, firms are not necessarily on their demand curves for labor nor are households necessarily on their supply curves.

[4] Let x_d' be the quantity of labor demanded per unit of capital in the private sector which produces goods purchased by both the private sector and the government. Then $w/A(\mathrm{t}) = F^*(x_d')$ or $x_d' = F(w/A(\mathrm{t}))$, where F is the inverse of F^*, is the labor demand function. The government, which is not a profit maximizer in the usual sense, demands x_d'' of labor per unit of capital in the economy. Then $x_d = x_d' + x_d'' = F(w/A(\mathrm{t})) + x_d''$. Since G_1 is defined as x_d'', then $x_d = F(w/A(\mathrm{t})) + G_1$. It is assumed that all of the capital is in the sector producing goods. If $G_1 = \xi G$, where G is the total demand by the government for goods and services per unit of capital, then, $x_d = F(w/A(\mathrm{t})) + \xi G$.

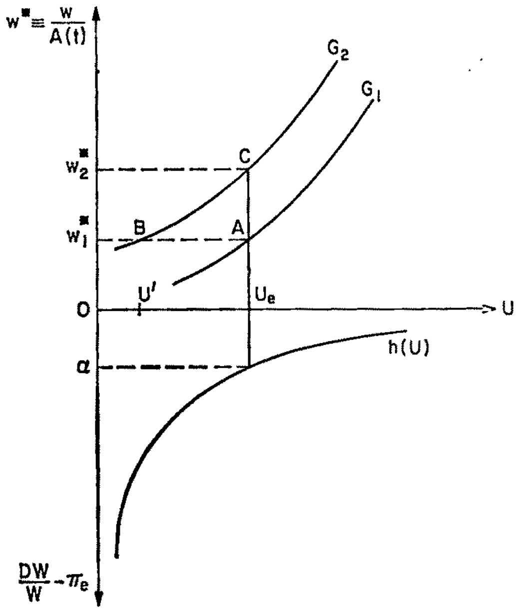


FIGURE 2. THE DETERMINATION OF THE EQUILIBRIUM UNEMPLOYMENT RATE U_e AND ADJUSTED REAL WAGE w^*

and services per unit of capital.[5]

$$(9) \qquad H^{-1}(U) = f\left(\frac{w}{A(t)}, G\right)$$

This basic equation is illustrated in the top half of Figure 2. The unemployment rate is positively related to the trend adjusted real wage $w/A(t)$, and negatively related to the government's direct demand for labor services.

Differentiate equation (9) with respect to time and derive equation (10). Although the coefficients β_1 and β_2 are functions of time, I treat them as relatively constant values in order to work with a time invariant linear system. Coefficient β_1 is an elasticity.

$$(10) \qquad DU = \beta_1\left(\frac{Dw}{w} - \alpha\right) - \beta_2 DG$$

where $\beta_1 \equiv f_1/(H^{-1})' \, w/A > 0$; $\beta_2 \equiv f_2/-(H^{-1})' > 0$ and $DA/A \equiv \alpha$. Substitute (6) into (10), since $DW/W - \pi = Dw/w$, and derive differential equation (11) for the change in the unemployment rate. If the linear form of the wage change equation (7) were used, then differential equation (12) would be derived.

$$(11) \qquad \begin{aligned} DU &= \beta_1(\pi^* + h(U) - \pi - \alpha) - \beta_2 DG \\ &= \beta_1 h(U) - \beta_1\pi + \beta_1\pi^* \\ &\quad - \alpha\beta_1 - \beta_2 DG \end{aligned}$$

$$(12) \qquad \begin{aligned} DU &= \beta_1(\pi^* + h_0 - h_1 U - \pi - \alpha) \\ &\quad - \beta_2 DG \\ &= (-\beta_1 h_1)U - \beta_1\pi + \beta_1\pi^* \\ &\quad + \beta_1(h_0 - \alpha) - \beta_2 DG \end{aligned}$$

These formulations imply Friedman's "natural" (equilibrium) rate of unemployment U_e because the coefficient of π^* in equation (1) is unity. When the actual rate of price change is fully anticipated, and both U and G are constant, then (12) implies (13).

$$(13) \qquad U_e = (h_0 - \alpha)/h_1$$

Equations (12) and (13) are macroeconomic relations which reflect the distribution of employment and unemployment among various sectors, each of which may have a different relation between the unemployment rate and the corresponding rate of nominal wage change. The dispersion of the unemployment may be reflected by h_0 and h_1, and the equilibrium rate of price change may affect the dispersion of unemployment and employment. Therefore, the equilibrium rate of price change may affect the equilibrium unemployment rate via vector (h_0, h_1, α) even though the coefficient of π in equation (1) is unity. It would be surprising if the macro-economic equilibrium rate of unemployment U_e were constant over a decade, since the distribution of employment and unemployment are unlikely to be constant. This is an empirical question to be examined. In-

[5] Since x_s is considered to be relatively constant, it will not be written explicitly in the $f(\cdot)$ function.

telligent economic policy must take into account the likelihood that (h_0, h_1, α) may not be constant over time, and may even be related to the equilibrium rate of price change.[6]

Define $x_1 \equiv U - U_e$, the deviation of the unemployment rate from the equilibrium rate U_e. Then $Dx_1 = DU$. Substitute (13) into (12) and derive equation (14), the differential equation for the unemployment rate.

$$(14) \quad Dx_1 = -\beta_1 h_1 \cdot x_1 - \beta_1 \pi + \beta_1 \pi^* - \beta_2 DG$$

This equation states that the *change* in the deviation x_1 between the actual and the equilibrium unemployment rate is negatively related to the deviation x_1, negatively related to unanticipated inflation $(\pi - \pi^*)$, and negatively related to the change in real purchases of goods and services by the government. Variable DG is a control variable, variables π and π^* are state variables which I shall now discuss.

B. *Price Change Equations*

A virtue of the widely used adaptive expectations equation (15) is that expectations are allowed to change slowly but smoothly in light of past experience. If the reader were asked to predict the rate of inflation over the coming year, the expected rate of monetary expansion would be an important variable in determining his forecast; and he would not confine his attention to a weighted average of past rates of price change. Further work must be directed to replacing this equation with something better.[7] In the interim, I use (15) to reflect expectations which may be changing slowly.

$$(15) \quad D\pi^* = b(\pi - \pi^*)$$

where $(\pi - \pi^*)$ is unanticipated inflation or deflation.[8]

The actual rate of price change π is the resultant of excess demand and cost pressures. If there were no excess demand for goods and services, the rate of price change would equal the growth of the nominal wage DW/W less the trend rate of technical progress α. There seems to be a consensus that this is a measure of cost pressure.[9] If $(DW/W - \alpha)$ were zero, then the rate of price change would be proportional to E, the excess demand for goods per unit of capital. Coefficient λ_p is a finite speed of response.[10]

$$(16) \quad \pi = DW/W - \alpha + \lambda_p E$$

My discussion of the determinants of the excess demand for goods per unit of capital

[6] This issue has been discussed by Tobin (1972a, b) and in the *Brookings Papers* in connection with the subject: Has the Phillips curve shifted?

[7] See Tobin (1972a) and the author, pp. 64–66.

[8] There are two extreme cases of equation (15). In (a), the expected rate of price change is equal to the current rate; in effect b is infinite.

$$(a) \quad \pi^* = \pi$$

Alternatively, b is zero, and the expected rate of price change is equal to the long-run steady-state value π_e. We know that π_e is equal to the rate of monetary expansion μ less the long-run growth of effective labor n.

$$(b) \quad \pi^* = \pi_e = \mu - n$$

[9] The price change equation used in much of the recent econometric literature (see Tobin (1972a)) is:

$$\pi = a_{12}\frac{DW}{W} - a_{13}\alpha + a_{14}\pi^* + f(U - U_e)$$

where U_e is the average or normal unemployment rate. In these empirical studies the unemployment rate is taken as exogenous. The main findings are that $a_{12} = a_{13} = 1$, and a modest effect of demand pressure is estimated, where $U - U_e$ is negatively related to demand pressure.

[10] I could have used the price change equation which characterized the "synthesis" model of money and growth (see the author, ch. 5): $\pi = \pi^* + \lambda_p E$, where π^* is the expected rate of price change. The rationale behind this equation is that specialists change prices on the basis of expectations π^* and market conditions E. Equation (16) in the text above implies that

$$\pi = \pi^* + h(U) - \alpha + \lambda_p E = \pi^* + \phi(E, U, \alpha)$$

which is formally similar to the price change equation in the synthesis model. See the author and Infante, pp. 539–41.

will be terse, since it is largely based upon my 1971 study.

C. *The Excess Demand for Goods*

The excess demand for goods per unit of capital E is equal to planned investment per unit of capital I/K plus planned consumption per unit of capital C/K plus real government purchases per unit of capital G less output per unit of capital Y/K. Define planned savings as output less planned consumption. It follows that

$$E = \frac{I}{K} - \frac{S}{K} + G \tag{17}$$

is the real excess demand for goods per unit of capital.[11]

Investment decisions are made by firms, and are independent of the savings decisions of households. The following model of the investment process was derived from Keynes' analysis in the *Treatise*, I, pp. 200–09. The desired change in the ratio of capital to output is positively related to the demand price of capital relative to its supply price. The supply price is just the price of a unit of current output $P(0)$. The demand price (capital value) is equal to the present value of the expected rents. Therefore, the desired change in the ratio of capital to output is positively related to the ratio of the present value of quasi rents to the price of a unit of current output. Let $R(\mathrm{t})$ be the expected marginal physical product of the unit of capital at time t; $P^*(\mathrm{t})$ the expected price of output at time t; $P(0)$ the current price of output; δ the depreciation rate; and ρ the expected constant nominal rate of interest equal to the current rate. Then the ratio of the demand price of capital to its supply price is given by (18).

[11] Alternatively, define E as the excess demand for capital: $K_D+C+G-(K_0+Y)$ where K_D is the stock desired and K_0 is the existing stock. Since $K_D-K_0=I$, the formulations are equivalent.

$$q = \frac{\text{demand price of capital}}{\text{supply price}} = \int_0^\infty \frac{R(\mathrm{t})P^*(\mathrm{t})}{P(0)} e^{-\rho \mathrm{t}} d\mathrm{t} \tag{18}$$

Assume that: (i) the price of output is expected to change at proportionate rate π^* such that $P^*(\mathrm{t})=P(0)e^{\pi^*\mathrm{t}}$, and (ii) the marginal physical product of capital $R(\mathrm{t})=re^{-\delta\mathrm{t}}$, where r is the current marginal physical product of capital. Equation (18) can be written as (19) when the denominator is positive; i.e., when the nominal rate of interest ρ exceeds the expected appreciation $\pi^*-\delta$ of the asset.

$$q = r\int_0^\infty \exp\left[-(\rho+\delta-\pi^*)\mathrm{t}\right]d\mathrm{t} = \frac{r}{\rho+\delta-\pi^*} \tag{19}$$

The desired change in the capital-output ratio is assumed to be positively related to $(q-1)$, the gap between the demand price of capital and its supply price. When $q>1$ or $r+\pi^*-\delta>\rho$, then firms wish to raise the ratio of capital to output. When $q<1$ or $r+\pi^*-\delta<\rho$, they wish to lower the ratio of capital to output.

When the demand price of capital is equal to the supply price, $q=1$ or $r+\pi^*-\delta=\rho$, then firms wish to maintain the existing ratio of capital to output. Suppose that effective labor were growing at proportionate rate n, and firms (correctly) expect output to grow at this rate in the long run. Then the desired growth of capital I/K is given by equation (20). This is the investment function.

$$I/K = n + g(r+\pi^*-\delta-\rho) + \delta \tag{20}$$

Investment function (20) consists of two parts: The first term refers to the equilibrium desired growth of capital, and the second term refers to the adjustment of the capital-output ratio to the desired

proportion. Speed of response g is assumed to be finite.[12]

Output per unit of capital y depends upon the ratio of effective labor per unit of capital x, as described by equation (21).

$$(21)\quad y = y(x); \qquad y' > 0, \quad y'' < 0$$

The production function is assumed to have the standard properties. The marginal product of capital r is derived from (21) in the usual manner.

$$(22)\quad r = r(x); \qquad r' > 0$$

Substitute equation (5) into these equations and derive equations (23) and (24), where output per unit of capital and the marginal product of capital depend upon the unemployment rate, given x_s.

$$(23)\quad y = y[(1-U)x_s] = Y(U); \quad Y' < 0$$

$$(24)\quad r = r[(1-U)x_s] = R(U); \quad R' < 0$$

A rise in the unemployment rate lowers $x=(1-U)x_s$ and thereby reduces the marginal product of capital. Physical capital is assumed to be fully employed at all times, as a result of a perfectly flexible rental price.

Savings are defined as output less planned consumption. The latter is a function, homogeneous of degree one, in capital, output, and real financial wealth. Usually, private financial wealth is assumed to be currency plus some fraction of the value of government interest bearing debt. Define θ as the ratio of private financial wealth to the money supply (M). Then real financial wealth per unit of capital is θm, where $m \equiv M/pK$. Open-market equations affect θ and hence it may be considered a control variable. The consumption function, described by (25), depends upon y and θm.

$$(25)\quad C/K = C(y, m, \theta);$$
$$1 > C_1 > 0, \quad C_2 > 0, \quad C_3 > 0$$

Fiscal policy in the form of a change in taxes less transfer payments is assumed to affect consumption insofar as it operates upon private wealth. This approach is based upon the assumption that consumption depends upon wealth rather than upon disposable income, and it obviates the need to build a tax rate schedule into the model. Implicit in this analysis is the government budget constraint. Taxes less transfers, plus changes in the federal interest-bearing plus non-interest-bearing debt are identically equal to the value of goods and services purchased by the government.

The savings function is therefore defined by equation (26). Substitute (5) into (26) and derive (27).

$$(26)\quad S/K = y - C(y, m, \theta)$$
$$= S(y, m, \theta); 1 > S_1 > 0, S_2 < 0, S_3 < 0$$

$$(27)\quad S/K = S[Y(U; x_s), m, \theta]$$

Using (20) and (27), the excess demand for goods equation (17) can be written as equation (28). The excess demand for goods per unit of capital is a function of U, π^*, m, ρ and of G, x_s, and n.

$$(28)\quad E = \frac{C}{K} + \frac{I}{K} + G - \frac{Y}{K}$$
$$= n + g[R(U; x_s) + \pi^* - \delta - \rho] + \delta$$
$$- S[Y(U; x_s), m, \theta] + G$$

The real net stock excess demand for bonds per unit of capital by the private sector depends upon: (i) the expected yield on capital $r+\pi^*$; (ii) the nominal rate of

[12] It is interesting to note that along an optimal growth path, the equation for $Dk = D(K/N)$, where capital is K and effective labor is N, is given by

$$Dk = -\frac{A(k)}{f''(k)}[f'(k) - (\lambda + \delta)]$$

where $f(k)$ is output per unit of effective labor, λ is the sum of the growth rate of effective labor and the depreciation rate, δ is the social discount rate, $A(k)$ is positive and $f''(k)$ is negative. See Infante and the author.

interest ρ; and (iii) real balances per unit of capital m. The first variable is an opportunity cost for owners of wealth, so that the excess demanded is negatively related to $R(U)+\pi^*$. For the usual reasons, the net excess demand for bonds by the private sector is positively related to the nominal rate of interest ρ. Real balances and real bonds are complementary assets in portfolios, so that the quantity of real bonds demanded per unit of capital should be positively related to real balances per unit of capital. Moreover, the quantity of real bonds supplied per unit of capital should be negatively related to the stocks of real balances per unit of capital held by firms. The real private excess demand for bonds per unit of capital can be written as $B(R(U)+\pi^*, \rho, m)$, where $B_1<0$, $B_2>0$, $B_3>0$.

Let z be the real stock of government interest-bearing debt per unit of capital. Then equilibrium in the bond market implies that:

$$B(R(U) + \pi^*, \rho, m) = z \tag{29}$$

Equation (30) describes the nominal rate of interest which equilibrates the bond market, and the partial derivatives have the usual signs.

$$\rho = \rho(U, \pi^*, m, z) \tag{30}$$

where $\rho_1=-B_1R'/B_2<0$, $\rho_2=-B_1/B_2>0$; $\rho_3=-B_3/B_2<0$; $\rho_4=1/B_2>0$. Since z is positively related to θm (real private wealth per unit of capital), the interest rate equation can also be written as (31) where g_i and ρ_i have the same signs. The important point is that, given m, a rise in θ (or z) raises the nominal rate of interest by shifting the traditional LM curve upwards.

Walras' law for stocks implies that, since the bond market is in equilibrium, the excess demand for real balances is equal to the excess supply of real capital.

$$\rho = g(U, \pi^*, m, \theta); \quad g_1 < 0,\ g_2 > 0,\ g_3 < 0,\ g_4 > 0 \tag{31}$$

Substitute (31) into (28) and derive equation (32) for the excess demand for goods per unit of capital when the bond market is in equilibrium.

$$E = E(U, \pi^*, m, G, \theta; x_s) \tag{32}$$

Function E does not contain ρ, the nominal rate of interest, as an argument. The sign of $\partial E/\partial\theta$ is the subject of controversy. A rise in θ, given m, will shift the LM curve upwards and raise the nominal rate of interest. Private investment will thereby be inhibited. On the other hand, the rise in θ may raise consumption demand (via equation (26)), and thereby shift the IS curve upwards. The net effect of a rise in θ upon E is not obvious on a priori grounds; and it is at the heart of the monetarist controversy.

D. *The Differential Equation for the Rate of Price Change*

The actual rate of price change is derived by substituting (32) and (7) into (16), and equation (33) is obtained.

$$\begin{aligned}\pi &= \pi^* + h_0 - h_1 U - \alpha \\ &\quad + \lambda_p E(U, \pi^*, m, G, \theta; x_s) \\ &= P(U, \pi^*, m, G, \theta; x_s)\end{aligned} \tag{33}$$

where

$$\begin{aligned}\partial\pi/\partial U &= P_1 = -h_1 + \lambda_p E_1 \\ \partial\pi/\partial\pi^* &= P_2 = 1 + \lambda_p E_2 \\ m(\partial\pi/\partial m) &= P_3 = m\lambda_p E_3 \\ \partial\pi/\partial G &= P_4 = \lambda_p E_4 \\ \partial\pi/\partial\theta &= P_5 = \lambda_p E_5\end{aligned}$$

Note that the unemployment rate U, and the expected rate of price change π^*, affect both the costs of producing output and the excess demand for goods per unit of capital. This equation differs from the price adjustment equations commonly used in empirical work primarily because

state variable m and control variables G and θ explicitly appear. Note also that P_3 is an elasticity, for a reason that will be apparent shortly.

The proportionate rate of change of real balances per unit of capital is given by equation (34).

$$(34) \qquad \frac{Dm}{m} = \mu - \pi - \frac{DK}{K}$$

where μ is DM/M the rate of monetary expansion and DK/K is the growth of capital. Since $x_s \equiv AN_s/K$ is assumed to be relatively constant in the short run and effective labor AN_s grows at rate n, equation (35) can be used in place of (34).

$$(35) \qquad \frac{Dm}{m} = \mu - \pi - n$$

The current rate of price change π in equation (33) depends upon m, which is the integral of past policies.

$$m(t) = m(0) \exp \int^{t} (\mu(\tau) - \pi(\tau) - n) d\tau$$

For this reason the economic system responds slowly to current monetary policy described by $\mu(\tau)$, in contrast to the models where consumption depends upon disposable income. Differentiate (33) with respect to time, and derive equation (36):

$$(36) \qquad D\pi = P_1 DU + P_2 D\pi^* + P_3 \frac{Dm}{m} + P_4 DG + P_5 D\theta$$

Substitute (35), (15), and (14) into (36) and derive differential equation (37). When the long-run (or full) system is considered, then (34) rather than (35) will be substituted into (36); and DK/K will replace n.

$$(37) \qquad \begin{aligned} D\pi = & -P_1\beta_1 h_1 x_1 + (P_2 b - P_1\beta_1 - P_3)\pi \\ & + (P_1\beta_1 - P_2 b)\pi^* \\ & + (P_4 - P_1\beta_2) DG + P_3(\mu - n) \\ & + P_5 D\theta \end{aligned}$$

The construction of the dynamical system, where the ratio of effective labor per unit of capital is relatively constant, is now complete. The model contains three differential equations in three state variables: U, π, and π^*. The long-run system is obtained by relaxing the assumption that x_s is relatively constant.[13]

II. The Short-Run Dynamical System and its Equilibrium Properties

There are three state variables in the above short-run macro-economic model: the deviation $x_1 \equiv U - U_e$ of the unemployment rate from its equilibrium level; the rate of price change π; and the expected rate of price change π^*. Control variables are: the change in real government purchases of goods and services[14] per unit of capital DG; the rate of monetary expansion per unit of effective labor $\mu - n$; and open-market operations and debt management policies of the treasury $D\theta$. Equation (38) describes the dynamic model in terms of a vector-matrix differential equation whose components are equations (14), (15), and (37). Let vector $X = (x_1, \pi, \pi^*)$ and $C = (DG, \mu - n, D\theta)$. Then equation (38) can be written compactly as (39).

A flow chart (Figure 3) is helpful in visualizing this system of differential

[13] An easy way to do it is to assume that effective labor *supplied* grows exogenously at rate n. Then

$$(a) \qquad Dx_s/x_s = n - DK/K$$

Outside the steady state, planned savings need not equal planned investment. Assume that the growth of capital is a linear combination of planned savings and planned investment, as described by equation (b).

$$(b) \qquad DK/K = aI/K + (1 - a)S/K, \qquad 1 \geqq a \geqq 0$$

The complete model can now be reduced to four differential equations in four state variables: U, π, π^*, and x_s; and the phenomena of unemployment and price changes are viewed in the context of a growing economy. See the author, Keizo Nagatani, and H. Rose.

[14] There is really another instrument that has been suppressed for the sake of simplicity: the distribution of government purchases between goods and direct labor services, ratio G_1/G. The taxes less transfer payments are implicit in the model via the government budget constraint.

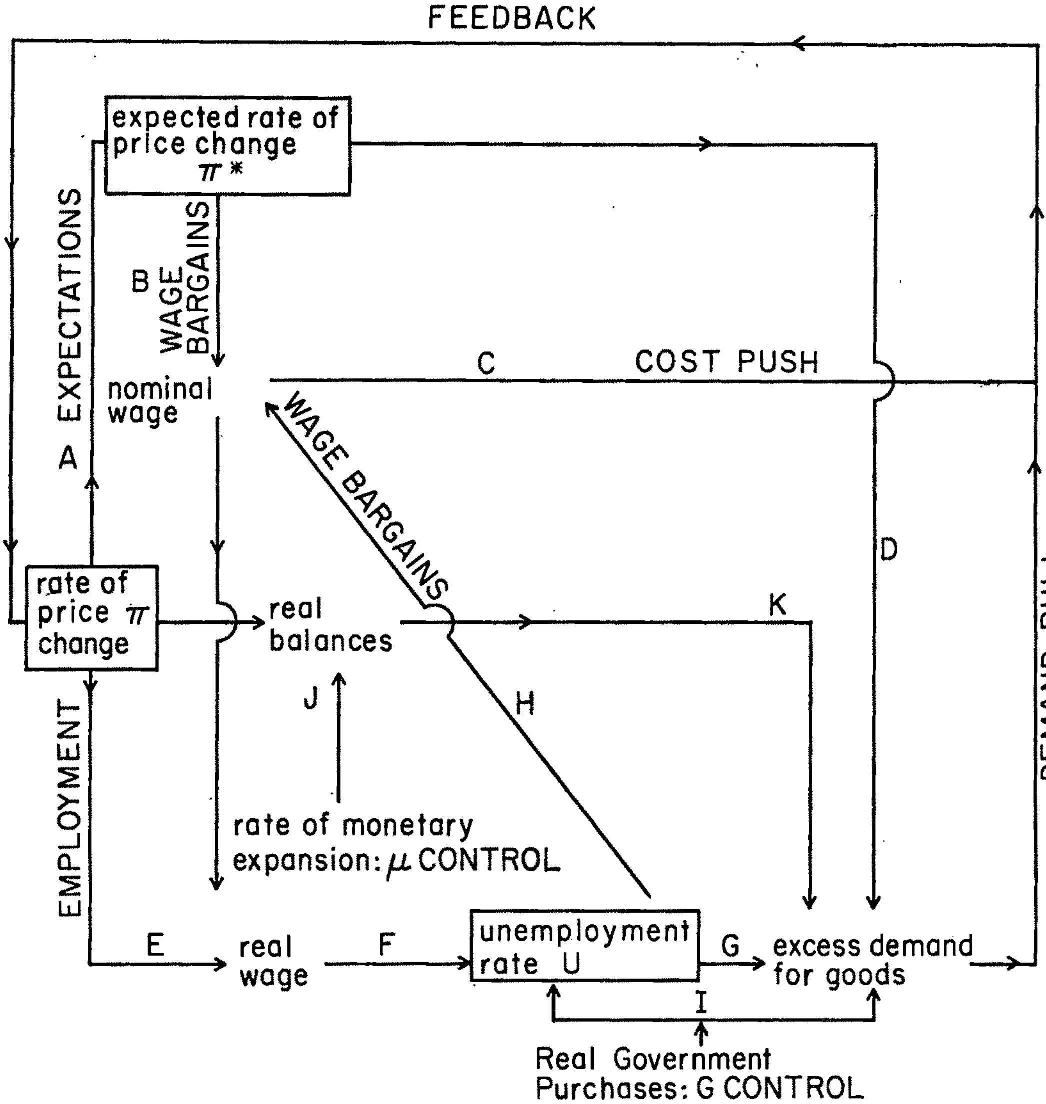


Figure 3. The Macro-Economic Model: Equation (38)

$$\begin{bmatrix} Dx_1 \\ D\pi \\ D\pi^* \end{bmatrix} = \begin{bmatrix} -\beta_1 h_1 & -\beta_1 & \beta_1 \\ -\beta_1 h_1 P_1 & (P_2 b - P_1\beta_1 - P_3) & (P_1\beta_1 - P_2 b) \\ 0 & b & -b \end{bmatrix} \begin{bmatrix} x_1 \\ \pi \\ \pi^* \end{bmatrix} + \begin{bmatrix} -\beta_2 & 0 & 0 \\ (P_4 - P_1\beta_2) & P_3 & P_5 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} DG \\ \mu - n \\ D\theta \end{bmatrix} \tag{38}$$

$$DX = AX + BC \tag{39}$$

equations. Wage change equation (6) is described by link B (price expectation term) and link H (the unemployment rate). Price change equation (33) is the sum of the cost push link C and demand pull link L. The adaptive expectations equation is link A. Price change expectations operate on prices directly through links B and C, and indirectly through link D, the excess demand for goods. The unemployment equation consists of links F and the left-hand side of link I.

Two policy inputs are depicted in the flow chart. Real government purchases of services operate directly upon the unemployment rate via the left-hand side of I; and real government purchases of goods operate upon the excess demand for goods, the right-hand side of link I. The second policy input is the rate of monetary expansion, link J. Changes in the rate of monetary expansion operate on money balances, and the excess demand for goods is affected via link K. Monetary policy can affect the unemployment rate indirectly by changing the real wage through link E. The circuitous route connecting monetary policy to the unemployment rate runs from J to K to E to F.

A. *Characteristics of the Steady State*

Define the steady state as a situation where the state variables X are constant ($DX=0$) and control variables DG and $D\theta$ are zero. It is *not* assumed that $\mu-n$ is zero. What are the characteristics of the steady state? Let subscript e denote steady-state values. Figure 2 is helpful in understanding the nature of the steady state. When the expected and actual rates of price change are equal, then (6) or (7) states that the growth of the real wage $(DW/W-\pi)_e$ is negatively related to the unemployment rate; and it is described by the curve in the lower part of Figure 2. When the unemployment rate U and real government purchases per unit of capital are constant, then according to (10), the real wage will rise at the trend rate of growth of productivity α. Only at unemployment rate U_e will the growth of the real wage $(DW/W-\pi)_e$ be equal to α. That is:

$$(DW/W - \pi)_e = \alpha = h(U_e) \tag{40}$$

Therefore, the equilibrium unemployment rate U_e depends upon the h function and the rate of technical progress α: the higher the rate of technical progress, the lower will be the equilibrium rate of unemployment.

The upper part of Figure 2 is based upon equation (9). It states that the unemployment rate is positively related to the adjusted real wage $w^*=w/A(t)$ and negatively related to the level of G. If there were a rise in G from G_1 to G_2, then the level of unemployment corresponding to w_1^* would decline from U_e to U'. However the real wage would rise at a faster rate than α if the level of unemployment fell below U_e. The rise in the real wage would lead to a decline in employment in the private sector; and the unemployment rate would rise above U'. Equilibrium would prevail when the real wage rose to w_2^* (point C). The transfer of labor to the labor-intensive G sector would raise the real wage, but no change would occur in the unemployment. The equilibrium adjusted wage $w_e^*=w/A(t)$ is given by equation (41). It is a function of U_e and G.

$$H^{-1}(U_e) = f(w_e^*, G) = H^{-1}[h^{-1}(\alpha)] \tag{41}$$

Two other important results occur in the steady state. There will be a zero excess deman for goods per unit of capital. This can be seen from (16), repeated here.

$$\pi = DW/W - \alpha + \lambda_p E \tag{16}$$

When the real wage grows at α, then

$$E(U, \pi^*, m; G, \theta, x_s) = 0 \tag{42}$$

The market for goods will be in equilibrium.

Finally, the equilibrium rate of price change π_e will be such that real balances per unit of capital will be constant. This follows from equation (36) or (37) when $x_1=\pi-\pi^*=DG=D\theta=0$. Then $P_3(\mu-n-\pi)=0$. Since P_3 is not zero,

$$\pi_e = \mu - n \tag{43}$$

The rate of price change is equal to the rate of monetary expansion less the growth of effective labor.

The excess demand for goods must be zero when the steady state is attained. Therefore,

$$E(U_e, \pi_e, m_e; G, \theta, x_s) = E(U_e, \mu - n, m_e; G, \theta, x_s) = 0$$

Solve this equation for m_e and obtain the equilibrium level of real balances per unit of capital. In this manner, the entire system is solved for its steady-state properties.

Assume that: (a) the Routh-Hurwitz conditions for dynamic stability are satisfied, and (b) each element along the principal diagonal of matrix A is negative. The second assumption states that if $(n-1)$ state variables are fixed at their equilibrium values, deviations of the nth variable from its equilibrium value will be eliminated asymptotically. The second assumption is convenient, but not necessary, for the subsequent dynamic analysis.

III. Unemployment and Inflation

A. *Analytical Techniques*

A graphic explanation of the unemployment cum inflation phenomenon can be presented through the use of phase diagrams, which are based upon vector-matrix differential equation (38).

Write the first equation in (38) as:

$$Dx_1 = (-\beta_1 h_1)x_1 - \beta_1\pi + \beta_1\pi^* - \beta_2 DG \tag{44}$$

When $\pi-\pi^*=DG=0$, then x_1 will converge asymptotically to zero; the unemployment rate will converge to the equilibrium U_e as described in (13). It is convenient, but not necessary, to assume that there is an equilibrium rate of unemployment which is independent of the rate of price change; i.e., that (h_0, h_1, α) in equation (13) are quite stable and are not significantly affected by macro-economic monetary and fiscal policy.

We know (Section IIA above) that $\pi_e^*=\pi_e=\mu-n$ in the steady state. To describe (38) in two dimensions so that a phase diagram can be used, it is useful to define a new variable z: the difference between the equilibrium rate of price change $\pi_e=\mu-n$ and current expectations. In the steady state, z will be zero.

$$z = (\mu - n) - \pi^* \quad \text{or} \quad \pi^* = \mu - n - z \tag{45}$$

Substitute (45) into the first equation of (38) and obtain (46).

$$Dx_1 = (-\beta_1 h_1)\cdot x_1 - \beta_1\pi + \beta_1(\mu - n) - \beta_1 z - \beta_2 DG \tag{46}$$

Define the EE' curve as the set of (x_1, π) along which $Dx_1=0$. It is equation (47) below.

$$\pi = -h_1 x_1 + \left[(\mu - n - z) - \frac{\beta_2}{\beta_1} DG\right] \tag{47}$$

This curve is negatively sloped. Why? If x_1 rises, the rate of growth of the nominal wage declines by $h_1\Delta x_1$ units. If the rate of price change declined by the same amount, the adjusted real wage $(w/A(t))$ would be constant; and hence the unemployment rate would not change. For this reason the EE' curve is negatively sloped.

Deviations of x_1 from the EE' curve, given π, tend to be eliminated because it was assumed that each element along the principal diagonal of (38) is negative, i.e., it was assumed that

$$\left.\frac{\partial Dx_1}{\partial x_1}\right|_{\pi} = -\beta_1 h_1 < 0 \tag{48}$$

The horizontal vectors in Figure 4 describe this phenomenon. What is happening in economic terms? If unemployment rate deviation x_1 is to the right of the EE' curve, the proportionate rate of change of the nominal wage will decline by h_1 units. Thereby, the growth of the adjusted real wage will decline by h_1 units. For every percentage point decline in the adjusted real wage, the employment rate will rise by β_1 units (equation (10)). It follows that deviations of the unemployment rate from equilibrium tend to be eliminated, given the values of the other state variables. Disequilibrium in the labor market ($x_1 \neq 0$) will produce repercussions upon the commodity market which, in turn, will affect the rate of price change; and thereby the labor market will be disturbed. Some of these effects will now be discussed.

Write the second equation in (38):

$$\begin{aligned} D\pi = &(-\beta_1 h_1 P_1)x_1 + (P_2 b - P_1\beta_1 - P_3)\pi \\ &+ (P_1\beta_1 - P_2 b)\pi^* \\ &+ [(P_4 - P_1\beta_2)DG + P_5 D\theta] \\ &+ P_3(\mu - n) \end{aligned}$$

Substitute $\pi^* = \mu - n - z$ into the above equation and derive:

$$\begin{aligned} D\pi = &(-\beta_1 h_1 P_1)x_1 \\ &+ (P_2 b - P_1\beta_1 - P_3)\pi \\ &+ (P_1\beta_1 + P_3 - P_2 b)(\mu - n) \\ &- (P_1\beta_1 - P_2 b)z + P_5 D\theta \\ &+ (P_4 - P_1\beta_2)DG \end{aligned} \tag{49}$$

Define the PP' curve as the set of (x_1, π) such that $D\pi = 0$. It is described by equation (50) and is drawn in Figure 4.

$$\pi = \frac{-\beta_1 h_1 P_1}{(P_3 + P_1\beta_1 - P_2 b)} x_1 + (\mu - n) + \frac{[(P_4 - P_1\beta_2)DG + P_5 D\theta - (P_1\beta_1 - P_2 b)z]}{(P_3 + P_1\beta_1 - P_2 b)} \tag{50}$$

The dynamic system is assumed to be stable; and the elements along the principal diagonal of matrix A are assumed to be negative such that inequality (51) is satisfied.

$$\left.\frac{\partial D\pi}{\partial \pi}\right|_{x_1} = -(P_1\beta_1 + P_3 - P_2 b) < 0 \tag{51}$$

The vertical vectors in Figure 4 describe this phenomenon. Given x_1, deviations in π tend to be eliminated. A unit rise in π above the PP' curve induces three distinct forces: some tend to produce further deviations of π, and others tend to restore it to its equilibrium value. These basic forces are reflected by $P_1\beta_1$, P_3, and $P_2 b$ and are described graphically in the flow chart (Figure 3).

(a) First: A rise in π raises the expected rate of price change by b units. Two subeffects are produced: one in the labor market and the other in the commodity market. The rise in the expected rate of price change raises the growth of the nominal wage; and the latter raises costs. Moreover, the rise in the expected rate of price change raises planned investment because the real rate of interest tends to decline relative to the rent per unit of capital. The increase in the excess demand for goods exacerbates the rate of inflation. As a result of a unit deviation of the rate of price change from equilibrium, the expected rate of price change rises by b units; and the rate of inflation is raised by $bP_2 = b(1 + \lambda_p E_2)$ units. This is an element of instability.

(b) Second: A rise in π tends to lower the growth of the adjusted real wage; and unemployment tends to decline by β_1 units. Several countervailing effects, summarized by $-\beta_1 P_1 = \beta_1(h_1 - \lambda_p E_1)$, are produced: (i) The growth of the nominal wage is increased by $\beta_1 h_1$ units and the resulting rise in costs is passed on in the form of a higher rate of price change. Inflation is aggravated. (ii) The decline in unemployment increases output; and both savings and

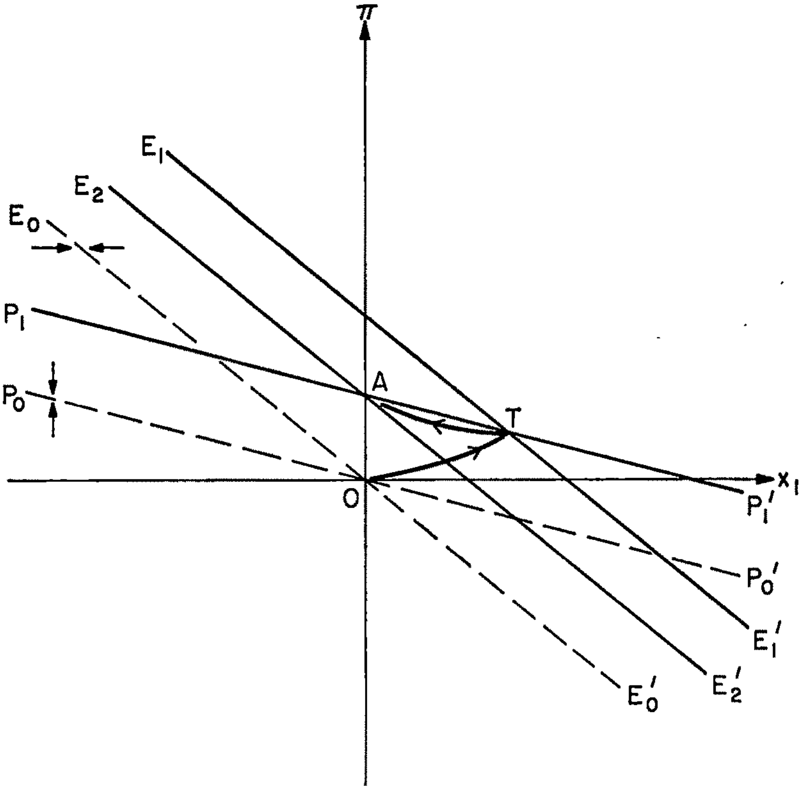


FIGURE 4. THE STEADY DECLINE IN $G(DG<0)$ AND THE RISE IN THE RATE OF MONETARY EXPANSION (μ) INDUCE THE ECONOMY TO FOLLOW TRAJECTORY OT. THERE IS INFLATION AND UNEMPLOYMENT. WHEN G LEVELS OFF, THE ECONOMY WILL HEAD BACK TO POINT A WHERE $(x_1, \pi)=(0, \mu-n)$

investment rise. The rise in savings tends to reduce the rate of inflation, whereas the rise in the rate of investment tends to increase the rate of inflation. It is not obvious on a priori grounds whether *excess* aggregate demand rises or declines; i.e., what is the sign of E_1? If the savings function is steeper than the investment function, when both are shown as functions of income, then E_1 is positive. *Excess* aggregate demand will decline; and the inflation will be mitigated.

(c) Third: A rise in the rate of price change produces a negative real balance effect. The real value of money and debt holdings is reduced. Aggregate demand is affected directly through consumption and indirectly through the liquidity preference function. The net effect is to reduce excess aggregate demand and lower the rate of price change. Term $-P_3=-m\lambda_p E_3$ is an element of stability.

It is assumed that the stabilizing forces predominate: $(P_3+P_1\beta_1-P_2b)>0$, and that the full Routh-Hurwitz conditions for stability are satisfied.

If the system is to be dynamically stable, then it is necessary that the slope of the PP' curve be algebraically greater than that of the EE' curve. The slope of the PP' curve will depend upon the sign of $-P_1=h_1-\lambda_p E_1$, an effect which was discussed above. In Figure 4, the PP' curve is drawn with a negative slope;[15] but the

[15] Preliminary empirical work suggests that the PP' curve is negatively sloped and that the system is dy-

analysis would be similar if it were positively sloped.

B. *The Phenomenon Explained*

It is convenient for graphic exposition (but not necessary for an algebraic analysis) to make a monetarist assumption (to be discussed in Section IV below) that $[(P_4-P_1\beta_2)DG+P_5D\theta]$, in equation (50) describing the PP' curve, is approximately zero. Moreover, assume that the numerator of the coefficient of z $(P_1\beta_1-P_2b)/(P_3+P_1\beta_1-P_2b)$ is sufficiently small as to be ignored.[16] Then the EE' and PP' curves can be described by equations (52) and (53), respectively.

$$(52)\quad \pi=-h_1x_1+\left[(\mu-n-z)-\frac{\beta_2}{\beta_1}DG\right]$$

$$(53)\quad \pi=\frac{-\beta_1h_1P_1}{P_3}x_1+(\mu-n)$$

When $\mu_0-n=z=DG=0$, the EE' curve will be described by E_0E_0' and the PP' curve will be described by P_0P_0' in Figure 4. The resulting equilibrium $(x_1, \pi)=(0, 0)$ is stable.

Suppose that real government purchases of goods and services per unit of capital were declining $(DG<0)$, but that the rate of monetary expansion increased from $\mu_0=n$ to μ_1. What will be the effects upon unemployment and inflation? The decline in G shifts the EE' curve (equation (52)) upwards from E_0E_0'. To maintain a given rate of unemployment, a rise in the rate of inflation is necessary to lower the growth of the real wage. In this manner the private sector would absorb the decline in the government's demand for labor services. Or given π, the unemployment rate increases when G declines. The new EE' curve that results when $DG<0$ but the forthcoming inflation is unanticipated $(z=\mu_1-n-0)$ has not been drawn in Figure 4; just keep in mind that it is above E_0E_0'.

The PP' curve (equation (53)) will be shifted up to the right from P_0P_0' to P_1P_1', as a result of the rise in $\mu-n$. Why? Given the rate of inflation, only a rise in the unemployment rate can offset the inflationary effect of rising real balances upon excess aggregate demand. Alternatively, real balances per unit of effective labor will be constant only if π rises to μ_1-n; hence, the PP' curve shifts upwards.

Initially the unemployment rate and rate of inflation rise. The latter increases because the growth of real balances raises the excess demand for goods by more than the greater softness in the labor market reduces the growth of the nominal wage. The sum of the two effects, $DG<0$ and the anticipated inflation $(z\rightarrow 0)$, shifts the EE' curve to E_1E_1'. When the inflation is anticipated, the E_1E_1' curve is: $\pi=-h_1x_1+(\mu-n)-(\beta_2/\beta_1)DG$. When $DG<0$, the intercept of E_1E_1' exceeds the intercept of the P_1P_1' curve.

Figure 4 describes the resulting situation. The economy will head along trajectory OT to point T. Both unemployment and inflation will increase as long as $DG<0$ and $\mu-n>0$. The paradox is easily explained within the context of this model.

When G stops declining and levels off $(DG=0)$, then the EE' curve will shift to E_2E_2'. The intercept will be at point $A=(0, \mu_1-n)$. As a consequence of the levelling off of G, the economy will head for point A in Figure 4 along a trajectory such as TA. If the speed of response in the labor market (h_1) is slow, it may take some time to reach equilibrium. The Keynesians question the ". . . ability [of an economy] to return automatically to full employment equilibrium within a reasonable

namically stable. This means that the conventional textbook case, where the savings function is steeper than the investment function, seems to be true.

[16] This assumption reduces the amount of curve shifting required for graphic exposition.

time (say, a year) if it is subject to the customary shocks and disturbances of a peacetime economy" (Patinkin, p. 901).

Clearly, a graphic description of a third order (3x3) system can only sketch its basic features. However, this model has been shown to be capable of explaining the qualitative features of the unemployment cum inflation paradox.

IV. The Monetarist Controversy

The dynamic model developed above can also explain the issues involved in the monetarist controversy. Thereby, the controversy about the roles of monetary and fiscal policies can be reduced to an econometric debate about empirical magnitudes.

The crucial propositions held by sophisticated monetarists can be briefly stated.

M1. "Changes in the pace of economic activity are not associated with any particular monetary growth and occur independently of the *level* of monetary growth. They only depend on the previous *changes* in monetary growth, whatever its inherited level" (Brunner, pp. 14–15).

M2. "The impact of monetary accelerations (or decelerations) on output and employment is essentially temporary" (Brunner, p. 13).

M3. "Monetary growth affects dominantly the price level" (Brunner, p. 13).

M4. "The slope of the *LM* curve is not the key difference between the monetarists and the neo-Keynesians" (Friedman, p. 910).

M5. "To have a significant impact on the economy, a tax increase must somehow affect monetary policy—the quantity of money and its rate of growth. Whether deficits produce inflation depends on how they are financed. If, as so often happens, they are financed by creating money, they unquestionably do produce inflationary pressure. If they are financed by borrowing from the public, at whatever interest rates are necessary, they may still exert some minor inflationary pressure. However, their major effect will be to make interest rates higher than they would otherwise be" (Friedman, p. 915).

M6. "Government spending unaccompanied by accommodative monetary expansion, that is financed by taxes or borrowing from the public, results in a crowding out of private expenditures with little, if any, net increase in total spending" (Andersen and Carlson, p. 8).

The monetarists could accept the differential equation for the unemployment rate:

$$Dx_1 = (-\beta_1 h_1)x_1 - \beta_1\pi + \beta_1\pi^* - \beta_2 DG$$

If the government does not hire labor directly, but just purchases goods from the private sector, then $G_1=0$. This is the usual case considered by the monetarists; however, I wish to consider the general case where the government purchases both goods and services.

The monetarists make the following assumptions:

(a) There is an equilibrium rate of unemployment which is independent of the rate of price change. It is their contention that (h_0, h_1, α) in equation (13) above are quite stable, and are not significantly affected by macro-economic policy. This assumption is necessary to derive monetarist conclusions. The EE' curve is equation (47), repeated here.

$$\pi = -h_1x_1 + \left[(\mu - n - z) - \frac{\beta_2}{\beta_1}DG\right] \tag{47}$$

$$\left.\frac{\partial Dx_1}{\partial x_1}\right|_\pi = -\beta_1 h_1 < 0 \tag{48}$$

(*b*) The expected rate of price change tends to adjust to the current rate with a

lag. Otherwise, monetary policy would not be able to exert even a temporary effect upon the unemployment rate. That is, if $\pi^*(t)=\pi(t)$ for all t, and $DG(t)=0$, then $x_1(t)$ will converge asymptotically to zero,[17] regardless of the rate of price change or the nature of current monetary policy.

(c) concerns the PP' curve. In general, the PP' curve is equation (50) repeated here.

$$\pi=\frac{-\beta_1 h_1 P_1}{(P_3+P_1\beta_1-P_2 b)}x_1+(\mu-n)+\frac{[(P_4-P_1\beta_2)DG+P_5 D\theta]}{(P_3+P_1\beta_1-P_2 b)}-\frac{(P_1\beta_1-P_2 b)z}{(P_3+P_1\beta_1-P_2 b)} \tag{50}$$

$$\left.\frac{\partial D\pi}{\partial\pi}\right|_{\pi 1}=-(P_1\beta_1+P_3-P_2 b)<0 \tag{51}$$

The monetarists argue that a rise in G, unaccompanied by a rise in the rate of monetary expansion, steadily raises the financial wealth-money ratio θ. Consequently the demand for real balances rises for two reasons: (i) A rise in G tends to raise the transactions demand for real balances directly. (ii) A steady rise in θ continues to raise the demand for real balances, because money and financial wealth are complementary assets in the portfolio. As a result of the rise in the demand for money, the LM curve keeps shifting upwards steadily and the nominal rate of interest rises and "crowds out" private investment. On balance the excess demand for goods does not change by much. Similarly, a decrease in taxes unaccompanied by a rise in the rate of monetary expansion raises wealth and induces an increase in spending. But it also increases the supply of securities or demand for money, and thereby raises interest rates. The latter crowds out private investment. The monetarist position may be interpreted to claim that in equation (50) the term in square brackets $[(P_4-P_1\beta_2)DG+P_5D\theta]$ is sufficiently small that it can be ignored.[18] This phenomenon has nothing to do with the slope of the LM curve but is concerned with its upward shift resulting from a rise in θ, the ratio of private financial wealth to the stock of money. Monetarists could write the price change equation as (54).

$$D\pi=(-\beta_1 h_1 P_1)x_1-(P_3+P_1\beta_1-P_2 b)\pi+(P_1\beta_1-P_2 b)\pi^*+P_3(\mu-n) \tag{54}$$

Again, it is convenient for graphic purposes to assume that the coefficient of z in (50) is small so that the PP' curve can be described by (55).

$$\pi=\frac{-\beta_1 h_1 P_1}{P_3}x_1+(\mu-n) \tag{55}$$

The EE' and PP' curves used in Section III above were based upon the monetarist model. To repeat, the *crucial monetarist assumptions* are that:

(a) Vector (h_0, h_1, α) is independent of monetary and fiscal policies, i.e., U_e is independent of these policies.

(b) Expectations change slowly.

[17] If $\pi^*(t)=\pi(t)$ for all t, then the first equation in (39) is: $Dx_1=-\beta_1 h_1 x_1$ or $x_1(t)=x_1(0)\exp(-\beta_1 h_1 t)$. It is independent of monetary shocks.

[18] The St. Louis monetarists support their contention on the basis of the following regression (Andersen and Carlson, p. 11):

$$\Delta Y(t)=2.67+\sum_{i=0}^{4}m_i\Delta M(t-i)+\sum_{i=0}^{4}e_i\Delta E(t-i)$$

$$\sum_{i=0}^{4}m_i=5.57\quad t=8.06;\qquad \sum_{i=0}^{4}e_i=.05\quad t=0.17$$

where $\Delta Y(t)$ is the dollar change in *GNP* in current prices in the t-th quarter; $\Delta M(t-i)$ is the dollar change in the money stock in quarter $(t-i)$; and $\Delta E(t-i)$ is the dollar change in high employment federal expenditures in quarter $(t-i)$. Alternatively, the monetarist position could be interpreted as claiming that P_5 is negative although the term in brackets is not negligible. Consequently, if G rose and then levelled off, the deficit financed by selling bonds would constantly raise θ. That is, DG becomes zero, but $D\theta$ remains positive. Therefore, deflationary pressure would be exerted ($D\pi<0$) which would offset the initial rise in G.

(c) Quantity $[(P_4-P_1\beta_2)DG+P_5D\theta]$ is approximately equal to zero so that government expenditures financed through the sale of debt do not exert a perceptible effect upon excess aggregate demand.

A diagrammatic explanation of the monetarist conclusions (M1–M4) can also be given on the basis of the above model summarized by equations (47), (48), (55), and (51).

Suppose that $\mu-n=DG=z=0$. Then the EE' and PP' curves are described by the EE_0' and P_0P_0' curves, respectively, in Figure 5. Both curves intersect at the steady-state solution, which is the origin in this case. The system is assumed to be stable. Directions of motion are described by the horizontal and vertical vectors; and the economy will converge to the steady state $x_1=0$, $\pi_e=\pi_e^*=\mu-n=0$. The trajectory is easily described on the basis of the phase diagram.

Assume that the economy is initially in equilibrium at the intersection of the E_0E_0' and P_0P_0' curves. Then let the rate of monetary expansion per effective worker accelerate from zero to $\mu-n=OA>0$ in Figure 5. If expectations change slowly, then the forthcoming inflation is not yet anticipated and $z=\mu-n-\pi^*=\mu-n-0=\mu-n$. Graphically, the EE' curve does not shift. Solely for the sake of geometric simplicity, assume that the coefficient of z in equation (50) is sufficiently small as to

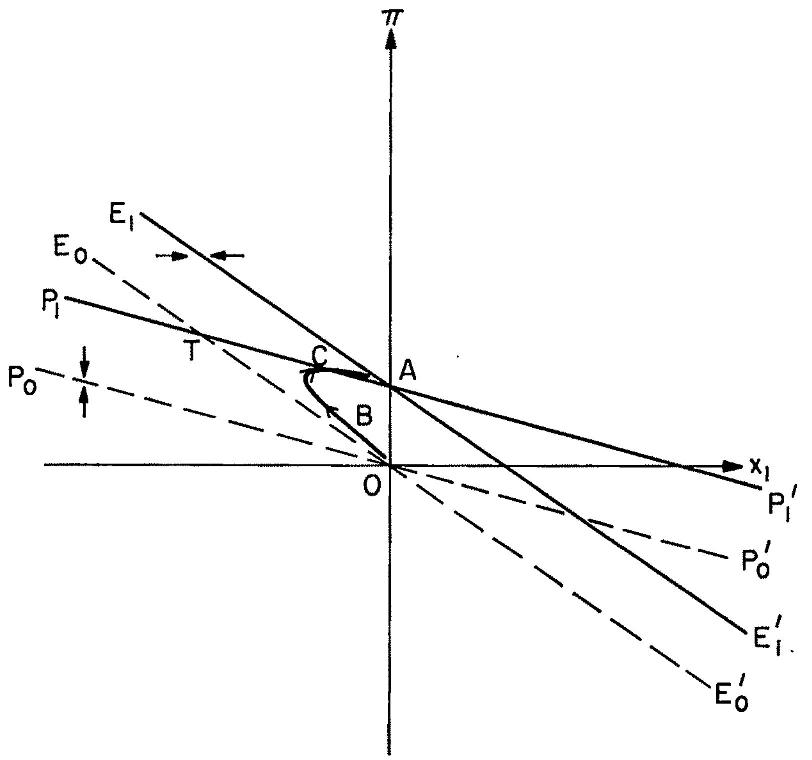


FIGURE 5. WHEN THE RATE OF MONETARY EXPANSION PER EFFECTIVE WORKER IS RAISED FROM ZERO TO OA, THE ECONOMY FOLLOWS TRAJECTORY $OBCA$ TO THE STEADY STATE $(x_1, \pi)=(0, \mu-n)$. GRAPHICALLY THE EE' CURVE SHIFTS ALONG P_1P_1' TO E_1E_1' (FROM E_0E_0')

be ignored.[19] Then the P_0P_0' curve will shift up by the full amount of the rise in the rate of monetary expansion.

Real balances per worker start to rise; and the excess aggregate demand is increased. The economy starts to head from the origin to the temporary equilibrium point T in Figure 5. Why? As a result of the rise in real balances, excess demand increases; and the rate of price change is raised. The inflation reduces the growth of the real wage, and thereby increases the rate of employment. For this reason, x_1 declines and π rises as the economy heads from the origin to point T.

As the inflation proceeds, inflationary expectations develop. This means that $z=\mu-n-\pi^*$ decreases. In graphic terms, this means that the E_0E_0' curve starts to rise towards E_1E_1', as wage negotiations are based upon a higher rate of anticipated inflation. The rising EE' curve implies that the economy will eventually head for point A, at the intersection of the P_1P_1' and E_1E_1' curves. Trajectory $OBCA$ may describe the motion of the economy to the new equilibrium point $A=(x_1, \pi)_e=(0, \mu-n)$.

Monetarist conclusions are thereby obtained. The monetary *acceleration* reduced the unemployment rate temporarily below the equilibrium rate. The absolute rate of monetary expansion per effective worker only affects the steady-state rate of price change; but there is no relation between unemployment rate x_1 and the constant rate of monetary expansion per worker at the steady state. Compare the origin with point A. Unemployment can be kept below U_e, i.e., x_1 can be kept negative, only if the process is repeated. That is, the rate of monetary expansion must be raised steadily; and this leads to hyperinflation. This model can imply monetarist conclusions.

If the term $[(P_4-P_1\beta_2)DG+P_5D\theta]$ were positive, then "weak monetarist" conclusions would follow. A steady rise in government expenditures $(DG>0)$ financed through the sale of bonds would raise the rate of price change and reduce the level of unemployment. The PP' curve would shift upwards and intersect the EE' curve in the second quadrant. However, if the level of government expenditures then stabilized $(DG=0)$ but $P_5<0$, then the crowding out effect would occur. The rising ratio of financial wealth to money would raise interest rates (the LM curve would shift upwards) and adversely affect the excess demand for goods. Graphically, the PP' curve would be shifted downwards when DG declines but $D\theta$ remains positive. Therefore, the monetarists would claim that government expenditures produce expansionary effects only if they are financed by the creation of new money.

A fiscalist on the other hand would claim that P_5 is positive. A rise in θ (say, resulting from a tax reduction where the deficit is financed through bond sales) would stimulate excess aggregate demand. To be sure the rise in θ, the ratio of bonds to money, would raise the nominal rate of interest by shifting the LM curve upwards. However, the rise in the debt would increase consumption demand, and shift the IS curve to the right. The net effect of a rise in θ will be expansionary if the IS curve shifts upwards by more than does the LM curve or if the IS curve is interest inelastic. Then P_5 will be positive. The monetarists would deny that this occurs.

I have shown what empirical specifications imply monetarist or fiscalist conclusions; and the controversy can be reduced to a set of testable propositions.

[19] The net effect is that, initially, the PP' curve shifts up by more than does the EE' curve. The assumption in the text describes this result with a minimum curve shift.

REFERENCES

L. C. Andersen and K. M. Carlson, "A Monetarist Model for Economic Stabilization,"

Fed. Reserve Bank St. Louis Rev., Apr. 1970, *52*, 7–71.

K. Brunner, "The Monetarist Revolution in Monetary Theory," *Weltwirtsch. Arch.*, 1970, Part 2, *105*, 1–30.

——— and A. Meltzer, "Friedman's Monetary Theory," *J. Polit. Econ.*, Sept./Oct. 1972, *80*, 951–77.

O. Eckstein, *The Econometrics of Price Determination Conference*, Washington 1972.

M. Friedman, "Comments on the Critics," *J. Polit. Econ.*, Sept./Oct. 1972, *80*, 906–50.

R. J. Gordon, "Inflation in Recession and Recovery," *Brookings Papers*, Washington 1971, *1*, 105–66.

E. Infante and J. L. Stein, "Optimal Growth with Robust Feedback Control," *Rev. Econ. Stud.*, Jan. 1973, *40*, 47–60.

J. M. Keynes, *A Treatise on Money*, Vol. I, London 1930.

D. Laidler, "Simultaneous Fluctuations in Prices and Output—A Business Cycle Approach," *Economica*, Feb. 1973, *40*, 60–72.

R. E. Lucas, Jr. and L. Rapping, "Real Wages, Employment and Inflation," *J. Polit. Econ.*, Sept./Oct. 1969, *77*, 721–54.

K. Nagatani, "A Monetary Growth Model with Variable Employment," *J. Money, Credit, Banking*, May 1969, *1*, 188–206.

D. Patinkin, "Friedman on the Quantity Theory and Keynesian Economics," *J. Polit. Econ.*, Sept./Oct. 1972, *80*, 883–905.

E. Phelps, "The 1972 Report of the President's Council of Economic Advisers: Economics and Government," *Amer. Econ. Rev.*, Sept. 1972, *62*, 533–39.

H. Rose, "Unemployment in a Theory of Growth," *Int. Econ. Rev.*, Sept. 1966, *7*, 260–82.

P. Samuelson, "The Role of Money in National Economic Policy," in *Controlling Monetary Aggregates*, Monetary Conference, Federal Reserve Bank of Boston, June 1969.

J. L. Stein, *"Money and Capacity Growth*, New York 1971.

——— and E. F. Infante, "Optimal Stabilization Paths," *J. Money, Credit, Banking*, Feb. 1973, Part II, *5*, 525–62.

J. Tobin, "Comments and Discussion: The Fellner, Okun, and Gordon Reports," *Brookings Papers*, Washington 1971, *2*, 511–14.

———, (1972a) "The Wage-Price Mechanism: Overview of the Conference," in O. Eckstein, ed., *The Econometrics of Price Determination Conference*, Washington 1972.

———, (1972b) "Inflation and Unemployment," *Amer. Econ. Rev.* Mar. 1972, *62*, 1–18.

———, (1972c) "Friedman's Theoretical Framework," *J. Polit. Econ.*, Sept./Oct. 1972, *80*, 852–63.

The Earnings and Promotion of Women Faculty

By George E. Johnson and Frank P. Stafford*

Given a widespread belief that women professionals experience discrimination, explicit public policy measures have been implemented to eliminate earnings and promotion differentials between men and women faculty. For example, Title IX of the Higher Education Act of 1972 states: "No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance." Yet neither the magnitude of the differential in earnings for given disciplines nor the extent to which factors other than discrimination are important determinants of such differentials were explored systematically prior to policy actions. The 1972 Act does allow "wage differentials based on a bona fide seniority or merit system," but where output is not readily quantified there is ample room for confusion and disagreement about "just" and "unjust" differentials. Implementation of public policy would be improved by a broader perspective on earnings differentials and by an explication of the role of factors other than discrimination.

In this paper we utilize data from the National Science Foundation (*NSF*) to examine: (i) the extent of the male-female salary and promotion differential for Ph.D.s in academic employment by field of specialization and, where data are sufficient, in business and government employment as well; and (ii) the extent to which lifetime choice of training and earning versus nonmarket work in the household influences the differential. We plan to develop the argument that the differential is smaller than commonly believed and that life cycle differences in labor force participation between men and women can be interpreted as the reason for a substantial share of the differential.

One distinct advantage to studying a rather well-defined labor market such as the academic labor market is that many sources of heterogeneity are removed; everyone in the analysis has a Ph.D., is in a given type of employment, and the like. Yet if females are systematically excluded from the occupation then discrimination can be thought of as taking place through entry limitations with women concentrated in selected (low paying, nonacademic) occupations.[1] Although we provide some tests of entry limitation for Ph.D. biologists, we are unable to answer difficult questions and allegations pertaining to the influence of such cultural phenomena

* University of Michigan. This research was sponsored by the National Science Foundation, GS-31639, and carried on in cooperation with the American Association of University Professors. Opinions expressed in this paper are ours. We would like to thank William Hershey for providing able assistance in this project. Operation with the large and varied data sets used in this paper was greatly facilitated by use of a computer software language, Micro, developed by the Labor Market Information Systems Project, directed by Malcolm Cohen.

[1] For one recent study which approaches the problem of discrimination against blacks from this perspective, see Barbara Bergmann. A related view has been advanced by sociologists, Robert Hodge and Patricia Hodge, ". . . discriminatory barriers may be raised to protect a majority of workers or even a controlling minority from competition from groups prepared to supply labor at lower costs" (p. 250).

as sexual stereotyping of occupational roles or the impact of such stereotyping combined with particular admission and recruitment policies. These and other factors, which can influence the number of women who do choose academic careers, are taken as given, and an explanation for the earnings differentials for those in academics is the object of this study.[2]

The plan of the paper is as follows: In Section I we review some implications of life cycle training theory concerning the experience-earnings profiles of men and women; Section II examines life cycle differences in academic compensation by sex and discipline; Section III presents additional analysis of life cycle differences; and Section IV provides a summary.

I. Life Cycle Training Theory and the Earnings of Women

Male-female differences in academic salaries are often attributed to discrimination; an alternative explanation is that the differential is primarily generated by the market's reaction to voluntary choices by females with regard to lifetime labor force participation and on-the-job training. In this section we will develop the implications of the latter hypothesis as a prelude to our analysis of the empirical evidence.

To explicate the life cycle differences in training and earnings of men and women faculty we can rely on human capital accumulation models which have been developed in recent years. These models typically derive paths of time allocation between training and earning, but it is also important to consider home productivity in order to contrast lifetime choices of men and women. The models can be divided into two classes depending on whether the (human) capital is general (equally productive in many firms), or specific (productive only in a particular firm). On the assumption that most of the human capital of academics is general, we first consider the implications of general human capital models. The most widely used of these consider the allocation of time between training and earning and do not consider leisure or home productivity. The objective function is simply discounted lifetime earnings. Consider the circumstances of individuals A and B, with identical earning capacity (stock of human capital) in the initial period, which is the point of completion of the Ph.D., and with equal ability to learn. We will assume that each individual has a time horizon of T years of potential future work experience, but, whereas A expects to remain in the labor force the full T years, B expects to drop out of the labor force between the years T′ and T″. In the model of Yoram Ben-Porath, the individual with the shorter work horizon would have a lower shadow price per unit of human capital and would thus choose less training at the outset of professional experience. This implies that the earning capacity of individual B would rise less precipitously than that of A, and this is reflected in the divergence in their stocks of human capital between time 0 and T′ in Figure 1. Between T′ and T″, B drops out of the labor force, so B's stock of human capital actually falls due to depreciation. After time T″, individual B reenters the labor force and again undertakes training, this time with the same

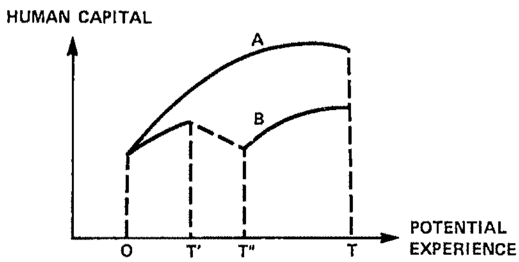


FIGURE 1

[2] Preliminary research on admission policies by graduate schools does not show substantial differential treatment of male and female applicants. See Lewis Solmon.

horizon as individual A. The time path of the actual earnings of the two individuals would reflect the price individuals must pay for training. For academics this price is embodied in the salary differential which exists between institutions with high and low training options[3]—for, given ability, starting faculty salaries are greater in the latter than in the former. In terms of our example, we would expect individual A to attempt to obtain a job at, say, Yale and individual B to accept a higher paying position at a less prestigious institution. Individual A would likely develop skills at a faster rate through his exposure to his distinguished senior colleagues, and these skills would command a high return in the future. Since individual B anticipates a shorter time in the labor force, it makes sense for him to forego the higher rate of skill improvement and collect a higher salary in the present.

The obvious application of the preceding discussion is to women. At least historically it has been common that a woman with some labor force experience during her lifetime ceases to participate or works on a part-time basis during at least a portion of the childbearing years. If a similar pattern of life cycle labor supply behavior applies to female academics, the average salaries of female relative to male academics would decline with years of potential experience up to the end of the childbearing years. When the youngest children reach school age, the woman who has dropped out reenters the labor force. The renewed employment, it has been argued by Jacob Mincer and Solomon Polachek "... is much more likely to represent a more permanent return to the labor force ... and strong incentives to resume investments in job related skills should reappear" (p. S-83). Therefore, later in their professional careers women faculty should tend to catch up to their male counterparts.

An alternative approach to explaining salary differentials is based on the assumption that, because of discrimination, females receive a lower rental price on their human capital. This, however, does not necessarily lead to the prediction that the earnings of females relative to males will decline over the life cycle. Although the returns to investing in an additional unit of human capital are lower for a group which is discriminated against, the opportunity costs of investment are also correspondingly lower.[4] Since the costs of market inputs (tuition, educational materials, etc.) are a very small portion of the total cost of on-the-job training, women would have the same incentives as men to undertake investments in skills—in the absence, that is, of expected time out of the labor force for home activities.

To some extent, human capital in academic departments is specific in nature—knowledge of programs, the dean's weaknesses, administrative procedures, and the like. In discussing her recent model of specific human capital Elisabeth Landes concludes: "It is clear ... that the existence of employer financed investment in [specific] human capital and differential turnover rates [by sex owing to home productivity differences] combine to be a sufficient condition for wage differentials to exist, even in the absence of any taste for discrimination by employers, coworkers, or customers" (p. 22). In academic life one activity that requires some specific training is administration. We will show that those engaging in administration,

[3] For an analysis of human capital accumulation through choices of jobs, see Sherwin Rosen.

[4] See Gary Becker, pp. 94–102. For a discussion of this point in the context of the Ben-Porath model, see Paula Stephan, pp. 51–53. It is often argued that discrimination causes a widening of the earnings gap between blacks and whites over the life cycle. However, some recent evidence shows that the apparent widening can be attributed to a vintage effect. See Finis Welch.

having a greater stock of specific human capital, have higher earnings. In line with Landes' model, women, having a higher (expected) turnover rate, would acquire less specific human capital and would have a lower probability of being assigned administrative tasks.

II. Life Cycle Differences in Academic Compensation by Sex and Discipline

A. *Lifetime Labor Force Participation*

The data source for our study is the *NSF* Register for 1970, and our analysis is restricted to Ph.D.s in particular academic disciplines on the presumption that they constitute, at least in the short run, separate labor markets. The fields which we included are anthropology, biology, economics, mathematics, physics, and sociology. The major variables of interest in our earnings function are experience (X), a measure of quality of graduate school attended ($RNKD$), whether noncitizen (NC), and whether female (FEM). From our discussion of the role of home productivity on lifetime earnings of women, it is clear that women's earnings should be expected to diverge from those of men because of family obligations, and this can be shown for women faculty.

The difference in the percent of respondents who report earning some income is essentially zero for single and married women during the first few years after receipt of Ph.D. but rises to about 15 percent at about 10 years after the degree and then returns to essentially zero at 15 and 20 years after the degree.[5] Our *NSF* data also show lower participation rates for females, and in Table 1 we have data for part-time employment of men and women in biology. Under age 30 part-time employment of those in the labor force is equally likely for men and women, and its level

TABLE 1—PERCENT OF MEN AND WOMEN PH.D. BIOLOGISTS EMPLOYED IN ACADEMIC INSTITUTIONS WORKING PART-TIME

	Age				
Sex	Under 30	30–34	35–44	45–54	55–64
Male	6.3	1.1	.5	.2	.4
(Sample Size)	(111)	(656)	(1453)	(1021)	(467)
Female	5.2	11.8	7.2	5.1	3.5
(Sample Size)	(58)	(229)	(390)	(315)	(149)

probably reflects job search of new graduates. In this group both men and women would very likely have experienced few delays in graduate study, and, further, for a woman under 30 to have completed her Ph.D., it is unlikely that she has put in time caring for small children. (Register data do not have information on marital status, children, or work history.)[6] At ages of 30 and older there appears a pattern of part-time employment of women which is very consistent with the presumed period of care for young (preschool) children. Eventually women biologists do reenter the labor force and their participation moves toward that of males, who work full-time throughout their careers. Although the number of women in other fields is too small to provide such an age-dependent comparison with males, the general results are the same.[7]

Turning to published data from the 1970 Census, one can obtain an estimate of differentials in hours worked per year by men and women in the occupational category of College and University Teacher. Multiplying an estimate of weeks worked per year times reported hours worked per week

[5] See National Academy of Sciences, p. 92.

[6] Given the nature of the *NSF* Register, we have had to proceed by extracting longitudinal inferences from cross-sectional data. New data sets based on panels would be highly desirable for subsequent research in this area. We do believe, however, that the inferences drawn here will remain valid when (and if) such panel data are available on a large-scale basis.

[7] For example, in 1970 mathematicians had the following percentages of full-time, part-time, and unemployed for men and women academics, respectively: 97.8, 1.2, 1.0; 86.9, 7.2, 5.9.

one can note that men put in about 1760 hours per year while women put in about 1200 hours per year.[8] This differential measures all hours including supplementary earnings and does not distinguish between Ph.D. and Masters degree holders. The general point it illustrates, though, is that even for women who do not drop out and interrupt their careers, there are fewer hours on the job and presumably a lower level of investment in on-the-job training.

B. *Functional Form of the Earnings Equation*

The earnings function employed in much of the human capital literature is a quadratic in years of experience.[9] From the *NSF* data we were able to construct estimates of potential years of postdegree experience (*XPO*) and predegree experience (*XPR*) through use of the respondent's reported year of highest degree (*DYR*) and years of professional experience (*X*).[10] Knowing that women faculty withdraw from full-time employment, we would then expect, from the models of general human capital, that the relative differential between male and female salaries would grow over time as measured by *XPO*. Accordingly, we specify a functional form which interacts our critical experience variables (XPO, XPO^2) with whether female (*FEM*).

From our previous research on the earnings of economists (see our 1974 paper) we know that the observed relation between earnings and postdegree experience varies systematically with the age at which individuals begin their professional experience. This is because a person with large amounts of predegree experience (*XPR*) is one who enters the profession with considerable delays and as a "late starter" will therefore have a lower shadow price for an additional unit of human capital. That is, he will have less time to capture the future gains from current on-the-job training, and we would expect him to be more likely to be priced out of the market for training options. Consequently, a late starter's lifetime earnings will be higher when he enters the profession ($XPO=0$) but should grow less rapidly with postdegree experience.[11] To capture this difference in on-the-job training we specify an interaction term ($XPO \times XPR$) in our earnings function. A final factor which we consider is the quality of the graduate program which the individual attended. This we approximate by whether the department was ranked in the top ten at the time the person attended (*RNKD*).[12] The dependent variable is the logarithm of nine-month salary, and our earnings function is fitted to those with the doctorate

[8] See U.S. Bureau of the Census, 1970. From Table 11, which gives weeks worked in 1969 for those who worked, one can obtain an estimate of weeks worked per year by sex (though not by age and sex). Multiplying this by an estimate of hours worked per week in 1970 given in Table 45 yields the estimated mean hours of 1760 for men and 1200 for women. The influence of familial activities on part-time employment for women doctorates is given in Helen Astin, p. 63.

[9] For studies which use a quadratic in experience to explain the logarithm of earnings, see Jacob Mincer or our 1973 paper.

[10] $XPO=1970-DYR$ and $XPR=X-XPO$.

[11] It could also be argued that those entering their careers late in life would be less likely to have received fellowship support (a rough measure of ability), and a shorter training period for less able persons is consistent with most life cycle models. See Stephan, pp. 126–27.

[12] Our procedure was as follows: Those respondents who received their doctorates prior to 1940 were assigned the ranking from the Hughes study of 1925; for graduates in the 1940–59 period, the rankings of the 1957 Keniston study; and for 1960–70 graduates, the 1964 Cartter ranking. In this way the individual's rank depends on both the school attended and his vintage. Those schools with top ten ranks were combined into a single dummy variable, *RNKD*. These data are found in A. M. Cartter, p. 35. This procedure had to be modified for biology since it is comprised of several disciplines which are ranked separately in the Cartter volume. Basically, the biology rankings were weighted averages of the separate subdisciplines in 1964 because historical data were not consistently available.

who worked full-time in 1970. The functional form specified is:

$$\begin{aligned}(1)\quad lnSAL = {} & \beta_0+\beta_1 XPO+\beta_2 XPO^2+\beta_3 XPR \\ & +\beta_4 XPR^2+\beta_5 XPO\text{x}XPR \\ & +\beta_6 NC+\beta_7 RNKD+\beta_8 FEM \\ & +\beta_9 FEM\text{x}XPO \\ & +\beta_{10} FEM\text{x}XPO^2 \\ & +\beta_{11} FEM\text{x}RNKD \\ & +\beta_{12} FEM\text{x}XPR \\ & +\beta_{13} FEM\text{x}XPR^2+\epsilon\end{aligned}$$

C. *Estimated Earnings Function*

In terms of our hypothesized role of voluntary choices of women regarding participation and earning, the critical expectations are that $\beta_9<0$ and $\beta_{10}>0$. Given appropriate magnitudes for these parameters, women's earnings will diverge from men's over some initial period of potential experience subsequent to completion of the degree, but their earnings will then converge back toward those of men when children enter school and they return to full-time employment and on-the-job training. Similar expectations would hold for β_{12} and β_{13} if predegree experience were also important in determining earnings and if women experienced interruptions in predegree experience during periods of child care. In four fields we performed Chow tests of the hypothesis that these two interaction terms add to the explanatory power of the model. We found the F-values to be weak and often insignificant even with the large sample sizes, which are given in Table 2.[13]

Excluding the *FEM* x *XPR* and *FEM* x XPR^2 variables, the estimated parameters of amended equation (1) are presented in Table 2. The results imply that persons with large amounts of predegree experience will have relatively flatter earnings profiles and for given total years of experience a person who has relatively more predegree experience will have a lower salary.[14] This bears on the earnings differentials between men and women because women have relatively larger amounts of predegree experience (see Table 4). For example, if total reported experience ($X=XPO+XPR$) and its square are used as the experience variables, then for 1970 the average percentage salary disadvantage typically increases by an average of about 2 percentage points over those reported in Table 3.[15] A life cycle training interpretation would be that given family obligations females take longer to complete their degrees and hence have profiles more akin to those of late starters. That is, their salary profiles are relatively flat. Alternatively, one could argue that if females receive less adequate fellowship support, one would also observe the same pattern because inadequate support tends to delay completion of the thesis. A recent study of the graduate student support of current Ph.D.s revealed that although women are more likely to receive no government support, they are more likely to receive major support from their universities. It was also found that women were somewhat more likely to receive postdoctoral fellowships (18 percent compared to 16 percent for men).[16]

[13] The F-statistics were: physics, .2; math, 5.1; economics, .5; anthropology, 3.0. (The critical F-value is 2.6 at the 5 percent level.) Thus, the model is only very weakly improved by specifying these interactions. (The signs were not of any consistent pattern either.) In contrast, interaction terms between *FEM* and *XPO* and *FEM* and *RNKD* yielded F-values of 5.2, 16.4, 3.8, and 5.5 in the respective fields.

[14] Namely, that β_1, $\beta_3>0$; $\beta_1>\beta_3$; β_2, $\beta_4<0$; $\beta_5<0$ and that $\beta_1+2\beta_2>\beta_3+2\beta_4$ over the range of 0 to 10 years of *XPO* and *XPR*. (Most Ph.D.s have less than ten years of predegree experience.)

[15] Using the interactive form with experience disaggregated, the differentials in predicted *starting* salaries ($XPO=0$) fall by a little over 1 percentage point if the profiles are evaluated for the respective mean values of *XPR* for females and males. This is because women tend to have more predegree experience.

[16] See National Academy of Sciences, pp. 76–82. See also Seymour Warkov, Alan Berger, and Bruce

TABLE 2—DETERMINANTS OF ACADEMIC SALARIES FOR SELECTED DISCIPLINES, 1970

Variable	Economics	Sociology	Anthropology	Mathematics	Biology	Physics
XPO	.0440	.0484	.0477	.0562	.0489	.0638
(β_1)	(.0013)	(.0023)	(.0027)	(.0011)	(.0014)	(.0015)
XPO^2	−.000763	−.000898	−.000724	−.000976	−.000831	−.001138
(β_2)	(.000037)	(.000068)	(.000079)	(.000029)	(.000039)	(.000042)
XPR	.0110	.0191	.0121	.0217	.0149	.0214
(β_3)	(.0021)	(.0030)	(.0045)	(.0015)	(.0022)	(.0029)
XPR^2	−.000203	−.000336	−.000082	−.000337	−.000138	−.000200
(β_4)	(.000090)	(.000119)	(.000210)	(.000066)	(.000110)	(.000146)
XPO x *XPR*	−.000677	−.001057	−.000564	−.001118	−.000763	−.000992
(β_5)	(.000099)	(.000151)	(.000224)	(.000076)	(.000100)	(.000142)
NC	−.036	−.071	.021	.023	−.105	−.025
(β_6)	(.016)	(.032)	(.027)	(.011)	(.016)	(.014)
RNKD	.060	.060	.0012	.021	−.009	.007
(β_7)	(.008)	(.012)	(.0142)	(.007)	(.009)	(.010)
FEM	−.054	−.042	−.088	−.063	−.119	−.094
(β_8)	(.032)	(.035)	(.038)	(.025)	(.019)	(.041)
FEM x *XPO*	−.0101	−.0096	−.0100	−.0171	−.0060	−.0200
(β_9)	(.0049)	(.0073)	(.0066)	(.0043)	(.0029)	(.0059)
FEM x XPO^2	.000228	.000196	.000066	.000258	.000057	.000495
(β_{10})	(.000147)	(.000287)	(.000206)	(.000124)	(.000088)	(.000157)
FEM x *RNKD*	−.089	−.061	−.066	.076	.013	.108
(β_{11})	(.036)	(.012)	(.040)	(.030)	(.021)	(.044)
Constant	4.714	4.589	4.597	4.573	4.515	4.465
(β_0)	(.010)	(.016)	(.019)	(.007)	(.010)	(.011)
R^2/*S.E.E.*	.415	.408	.619	.530	.475	.635
	.217	.223	.171	.203	.222	.192
Sample Size	3040	1659	672	5335	4059	1939

TABLE 3—ESTIMATED FEMALE-MALE RELATIVE SALARY BY YEARS OF EXPERIENCE BY DISCIPLINE, 1970[a]

Years of Postdegree Experience	Economics	Sociology	Anthropology	Mathematics	Biology	Physics
0	.947	.959	.916	.939	.888	.935
	(.031)	(.035)	(.036)	(.024)	(.017)	(.037)
5	.906	.918	.873	.868	.863	.864
	(.019)	(.021)	(.019)	(.014)	(.010)	(.020)
10	.876	.888	.834	.812	.841	.816
	(.020)	(.026)	(.021)	(.016)	(.010)	(.020)
15	.857	.868	.800	.770	.822	.789
	(.024)	(.030)	(.025)	(.019)	(.012)	(.024)
20	.848	.856	.770	.739	.806	.780
	(.027)	(.036)	(.026)	(.020)	(.012)	(.027)
25	.849	.852	.743	.719	.792	.786
	(.032)	(.055)	(.028)	(.021)	(.013)	(.027)
30	.859	.857	.720	.709	.781	.811
	(.042)	(.092)	(.039)	(.023)	(.017)	(.030)

[a] Estimated differences are given for individuals from unranked graduate schools and $XPR=4$, the approximate mean of predegree experience across the sample. Standard errors of the estimated values of $\log SAL_F - \log SAL_M$ are given in parentheses beneath the estimated relative salaries. Letting α_1, α_2, and α_3 be the coefficients on *FEM*, *FEM* x *XPO*, and *FEM* x XPO^2, respectively, this standard error is calculated according to the formula

$$s = (V(\hat{\alpha}_1) + XPO^2 v(\hat{\alpha}_2) + XPO^4 v(\hat{\alpha}_3) + 2\,XPO\,C(\hat{\alpha}_1,\hat{\alpha}_2) + 2\,XPO^2\,C(\hat{\alpha}_1,\hat{\alpha}_3) + 2\,XPO^3\,C(\hat{\alpha}_2,\hat{\alpha}_3))^{1/2}$$

TABLE 4—MEANS OF SELECTED VARIABLES BY SEX AND DISCIPLINE, 1970

	Economics		Sociology		Anthropology		Mathematics		Biology		Physics	
Variable	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Gross Professional Income	20,810	15,960	18,590	14,540	17,890	14,000	18,580	14,460	17,520	13,980	17,940	14,600
XPO	9.8	9.0	9.8	6.9	9.9	10.0	8.7	9.5	11.7	11.1	10.1	12.1
XPR	4.3	5.0	4.5	5.4	3.0	2.9	3.9	5.7	2.6	3.4	2.4	2.5
RNKD[a]	34.1	27.2	35.0	33.9	37.2	25.0	26.1	24.2	31.4	31.5	29.1	30.0
Sample Size	2893	147	1476	183	577	95	5079	256	3191	868[b]	1839	100[b]

[a] Percentage from graduate schools ranked in top ten.
[b] In biology and physics, females were sampled at a 100 percent rate and males at a 33 percent rate.

Examining the male-female salary differentials in Table 3, there are three notable features in the context of the life cycle training model: 1) at $XPO=0$ there should be the least obsolescence of skills for females. Here the differential is correspondingly the smallest, ranging from about 4 percent in sociology to 11 percent in biology; 2) the differential grows most rapidly over the years 5–15, the years when child care is most prevalent; 3) there is typically a narrowing of the female-male salary differential at advanced years of postdegree experience. If women return to full-time employment after child rearing years (as demonstrated by Table 1) and consequently reacquire skills, then the differential should cease to rise or even narrow as observed in our data. The rather limited number of women in the samples at these advanced points in their careers means that the standard errors on the ratio are high but the point estimates are consistent with the on-the-job training interpretation. That is, the differential either grows at a much lower rate (anthropology,

Frisbie, p. 79. They show that married women with children receive fewer stipends, but single women have stipend support comparable to men's.

TABLE 5—PREDICTED NINE-MONTH ACADEMIC SALARIES FOR MALES AND FEMALES IN SELECTED DISCIPLINES, 1970[a]

		Years of Postdegree Experience			
Discipline		0	10	20	30
Economics	Male	11,610	16,260	19,550	20,170
	Female	11,000	14,250	16,580	17,330
Sociology	Male	10,560	15,020	17,840	17,720
	Female	10,130	13,340	15,270	15,190
Anthropology	Male	10,400	15,230	19,310	21,180
	Female	9,520	12,710	14,870	15,250
Mathematics	Male	10,510	15,980	20,010	20,600
	Female	9,860	12,980	14,790	14,610
Biology	Male	9,680	14,090	17,360	18,130
	Female	8,590	11,850	13,990	14,150
Physics	Male	9,440	15,320	19,810	20,400
	Female	8,590	12,000	14,740	15,910

[a] Estimated differences are for individuals from unranked graduate schools with $XRP=4$.

mathematics, biology) or narrows (economics, sociology, physics).[17] The coefficients on the interaction terms between *FEM* and *RNKD* (β_{11}) imply that except for mathematics and physics, females who received their doctorates from the relatively prestigious schools do not have significantly higher salaries than females who attended other graduate schools ($\beta_7+\beta_9 \lesseqgtr 0$). This is not surprising in the light of the facts that (a) the major monetary benefits due to receipt of a doctorate from a prestigious institution accrue after large amounts of postdegree experience,[18] and (b) postdegree experience is not as consistently acquired for females as for males.

D. *Academic Rank of Men and Women and the Effect of Marriage*

Another index of total compensation (salary, fringe benefits, emoluments, recognition) is academic rank. If life cycle influences operate to generate a differential between men's and women's salaries, they should also operate to generate a differential in promotion.[19]

The differentials in the cumulative probability of attaining a rank of full professor (or dean) are shown for economics and biology in Table 6. Here the promotion

TABLE 6—PREDICTED CUMULATIVE PROBABILITY OF ATTAINING THE RANK OF FULL PROFESSOR OR DEAN FOR MEN AND WOMEN, 1970[a]

	Economics		Biology	
XPO	Males	Females	Males	Females
0–3	.046	.004	.054	.043
4–6	.167	.077	.098	.103
7–9	.451	.216	.186	.106
10–12	.734	.419	.417	.195
13–15	.854	.400	.664	.270
16–19	.929	.625	.806	.393
20+	.970	.539	.928	.619

[a] The functional form was:

$$TEN = \alpha_0 + \alpha_1 FEM + \alpha_2 XPR + \alpha_3 FEM(XPR) + \sum_{i=1}^{m} (\beta_i XPO_i + \gamma_i FEM(XPO_i))$$

where XPO_i are the intervals of postdegree experience given in Table 8 and $TEN=1$ if full professor or dean; 0, otherwise. In the calculations the assumed value of $XPR=5$.

pattern results in a cumulative differential which is built up in the years $XPO=4$–15 in economics and $XPO=7$–19 in biology.[20] Under the lifetime training view, the differential would not be expected to prevail if it were possible to isolate the effects of family obligations, and one way to do so would be to study single women. While our data do not contain marital status, a recent study of Ph.D. cohorts does. For those in a cohort centered in 1940 who remained in academic employment, about 81 percent had attained rank of professor twenty years later. For the married women in the cohort 46 percent had attained the rank of full professor twenty years later whereas for single women 70 percent had attained the rank of full professor twenty years later.[21] Adjusting for sample sizes,[22]

[17] We also analyzed the 1964 Register data. (Anthropology data were not available in 1964.) The same results 1)–3) obtained. Between 1964 and 1970 there were no substantial changes with the exception of economics where the differential fell at all levels of $XPO \geq 5$.

[18] See our 1974 paper for a development of the fanning out of profiles with years of postdegree experience between graduates of top ranked and other graduate programs.

[19] One approach would be to examine the male-female salary differentials within given academic ranks on the assumption that one must compare people at a given job level. Doing so by introducing academic rank into the equations of Table 2 reduces the male-female differential by over 50 percent. But this would be a poor specification, because rank and salary are rather closely tied. Moreover, it does not resolve the issue of the relative importance of acquired skill differences and discrimination in effecting a differential. This point has been discussed by Malkiel and Malkiel, p. 704.

[20] As an aside, one reason for the more rapid promotion in economics for both men and women could be a rank specific salary structure in some schools which requires that persons in higher paying fields be promoted early in order to retain them.

[21] See National Academy of Sciences, pp. 21, 71, 85. Single is defined as never married and married is defined as ever married.

[22] Males, 1958; single women, 83; married women,

the percent full professor is 85 for males. Thus the figures for Ph.D.s as a whole appear generally comparable to ours for economics and biology and reveal that about three-fifths of the promotion differential can be reduced by comparing single women rather than married women to men.[23]

E. *Differences in Specific Human Capital*

Skills specialized to a particular employer also constitute a form of human capital. If higher turnover of women is expected, both women and employers will have incentive to invest in smaller amounts of specific human capital. Administrative tasks in a given department often require particular abilities and knowledge which constitute specific human capital. Utilizing our sample of Ph.D. economists, we found that 23 percent of the men and only 8 percent of women reported administration as a major work activity.

In contrast to the specific human capital hypothesis, it can be argued that women are less frequently engaged in administration because they know that development of such skills will not be rewarded on a basis comparable to that of men so trained. Our analysis shows that administrative skills add to one's earning power and, further, women who do administer realize an earning increment roughly comparable to that for men. While administration substantially increases the nine-month salary of males (by 6.2 percent $\pm$ .9) it also increases the nine-month salary of females (by 10.5 percent $\pm$ 5.2).

III. Further Analysis

A. *Institution of Employment as a Measure of On-the-Job Training*

In our previous discussion, the argument was advanced that women would choose to acquire less training upon entry into the labor force. In the academic labor market, this choice is often reflected in the type of school of employment. For example, we find that the prestigious small private colleges hire their faculty from the highly ranked graduate schools in proportions equal to those hired by the large graduate programs (see Table 7, line 8). Yet, given roughly comparable graduate background, the experience-salary profile is more pronounced at the major graduate schools and rather flat at schools which do not emphasize research such as the prestigious small private schools. This differential profile shape can be attributed to differential investment in professional development, particularly given the professional rewards to research skills. That is, starting salaries in the schools which emphasize research are often lower but surpass the earnings in other schools for higher levels of experience. This initial salary differential can be interpreted as indexing the implicit price of additional training options. It should also be true that women are less likely to purchase such options for reasons already discussed. If women are therefore concentrated in schools which emphasize teaching, then we will have some understanding of the institutional relations which describe their flatter earnings profiles.

Utilizing our sample of mathematicians (1970), we created five subsamples by school of current employment: (i) the top twenty schools in the 1964 Cartter ranking; (ii) a selection of thirty-four highly

138. Given these small totals, disaggregation of the data by discipline would be difficult. See National Academy of Sciences, p. 72.

[23] There is evidence that single women are less distinguished academically than their married counterparts. See Lindsey R. Harmon. This could be interpreted as suggesting that if single women were as qualified as their married counterparts the (male-never married female) differential in promotion would be lower. However, it is also well known that women have better records than their male counterparts.

TABLE 7—MEANS OF VARIABLES IN 1970 SUBSAMPLES OF MATHEMATICIANS BY INSTITUTION OF EMPLOYMENT

Variable	Entire Sample	Top Twenty	Small Private	Great Plains	University of California	California State
SAL	15,380	15,670	13,590	14,290	15,460	14,570
INC	18,390	20,390	17,300	17,170	19,820	17,530
RES	.69	.85	.69	.64	.90	.50
XPO	8.7	9.8	9.1	8.2	10.1	9.6
XPR	4.0	2.5	2.6	4.9	2.3	5.2
FEM	.05	.02	.06	.05	.02	.10
NC	.07	.10	.04	.05	.07	.05
RNKD	.26	.49	.46	.09	.49	.29
$\frac{SAL\ (XPO=25)}{SAL\ (XPO=0)}$	1.90	2.32	1.89	1.79	2.39	1.72
Sample Size	5,335	728	154	274	201	119

rated small private colleges;[24] (iii) all colleges and universities in a collection of states on the Great Plains; (iv) all schools in the University of California system; and (v) all schools in the California state college system. Regressing the logarithm of salary on our experience variables, whether noncitizen, whether graduate school attended was ranked in the top ten, and a single 1–0 variable for whether female, we found smaller sex differentials within the school types than for the entire sample.[25] A single exception to this was the Great Plains subsample which contains quite a heterogeneous group of schools. At the same time, the proportion of females in each subsample varies from .02 of the top twenty and University of California subsamples to .10 in the California state subsample (Table 7). This implies that, for whatever reasons, women mathematicians are employed in institutions which do not specialize in research and do not offer relatively large salaries for senior faculty, and this result is consistent with our previously reported widening of the male-female salary differential. The conclusion that women are found in greater proportion only in schools where the major focus is not research was also found in the other two disciplines we studied in this fashion: economics[26] and biology. Thus women's flatter salary profile can be attributed to their employment in schools which place less emphasis on research and more on teaching.[27] For additional evidence of this, in biology, where the sample size for women is larger, the interactive form (*FEM*, *FEM* x *XPO*, *FEM* x XPO^2) was used. The respective female-male earnings ratios at the top twenty Cartter schools were .875, .755, and .725 at 0, 15, and 25 years of postdegree experience and were (for a weighted average) .925, .835, and .825 at the other school types.[28] An explanation of this different life cycle pattern across school types by sole reliance on the discrimination hypothesis seems to require the argument that faculty at the large

[24] Amherst, Antioch, Bates, Beloit, Bowdoin, Bryn Mawr, Carleton, Colgate, Connecticut College, Davidson, Dennison, Furman, Goucher, Hamilton, Haverford, Harpur, Kalamazoo, Middlebury, Mills, Mt. Holyoke, Oberlin, Occidential, Reed, Rollins, Sarah Lawrence, Smith, Stephens, Swarthmore, Trinity, Wellesley, Wesleyan, Williams, Wooster, and Vassar.

[25] A single shift variable (1–0, whether female) was required given the small sample sizes for women.

[26] See our 1974 paper.

[27] These results are consistent with those of Astin, p. 73. She demonstrates that about 40 percent of men Ph.D.s' time is spent in research in comparison to 25 percent of women Ph.D.s' time.

[28] Excluding the University of California schools, some of which are in the top twenty.

graduate departments have a *greater* preference for discrimination against women.

B. *Earnings Differences Within a Specific Institution*

There has recently been a tendency for individual colleges and universities to publish (or have published) all data on academic salaries, and this has motivated analyses of salary levels at individual schools with the question of sex discrimination in mind. The interactive model employed in this paper should be applied in such cases.[29]

For example, data on salaries and experience for faculty members at one university are published in *Michigan State University Salary Analysis*. Taking all individuals at or above the rank of assistant professor in the six fields in our study, we obtained the following result:

$$
\begin{aligned}
(2)\quad \ln SAL = {} & \underset{(.04)}{4.75} + \underset{(.003)}{.054X} - \underset{(.0001)}{.0011X^2} \\
& - \underset{(.111)}{.030FEM} - \underset{(.017)}{.026FEM \times X} \\
& + \underset{(.000471)}{.000894FEM \times X^2} \\
& + \underset{(.049)}{.130ECON} - \underset{(.098)}{.053ANTHR} \\
& - \underset{(.043)}{.065PHYS} - \underset{(.041)}{.058MATH} \\
& - \underset{(.037)}{.096BIO}
\end{aligned}
$$

where $R^2=.54$, $N=323$, and the coefficients on the five discipline variables represent deviations from the excluded group, sociology. These results suggest that at Michigan State University starting salaries for female academics are only 3 percent lower than for males (and not significantly so); the disadvantage of females grows to 20 percent (± 7) at 15 years of reported experience and then declines for higher levels of experience. We believe this qualitative result would be observed for any university in the United States for which the sample size is sufficiently large.

C. *The Earnings of Women and Noncitizens*

If (male) academics are specified as prone to sex discrimination, it is possible that discrimination against groups other than women would occur. In our sample noncitizens constitute such a group, and clearly they do not have a particular nonmarket role as do women. Using the interactive functional form, we find the life cycle earnings differentials for noncitizens relative to citizens decline over time.

As presented in Table 8, it is clear that the noncitizen-citizen profiles converge as would be expected if there are adjustment costs associated with immigration. If only discrimination were operating and both women and noncitizens were subject to the effects of this preference, we would then expect similar life cycle patterns. The pattern in Table 8 is consistent only with the view that women differ from noncitizens in family roles or that academics discriminate against one group (women) and not against another (noncitizens). Parallel results obtained in the other field which we studied in this respect (economics).

D. *Employment of Women Within Biology*

Are women overrepresented in lower paying subspecialties? If women's profile shapes can be described by knowing the schools in which they teach, can their systematically lower salaries also be related to their disciplines and subdisciplines? From our six fields it is clear that the percent of women in a field appears to be related to earnings level. Mathematics, economics, and physics have a lower percentage of

[29] See, for example, David Katz. Katz, however, does not employ an interactive model and thus only computes an average discrimination coefficient.

TABLE 8—NONCITIZEN-CITIZEN AND FEMALE-MALE LIFE CYCLE SALARY DIFFERENTIALS, TOP TWENTY GRADUATE SCHOOLS IN BIOLOGY, 1970[a]

	Years of Postdegree Experience					
	0	5	10	15	20	25
Noncitizen-Citizen	.772	.860	.978	1.047	1.053	.998
	(.060)[b]	(.033)	(.041)	(.063)	(.079)	(.095)
Female-Male	.876	.822	.782	.753	.734	.726
	(.040)	(.021)	(.018)	(.018)	(.024)	(.026)

[a] Assumed value of $XPR=5$.
[b] Standard errors in parentheses.

TABLE 9—PERCENT FEMALE AND EARNINGS FOR THE BIOLOGICAL SCIENCES, 1970

Discipline	Percent Female[a]	Gross Professional Income	$\partial ln\, SAL/\partial\, XPO$ $(XPO=0)$	Percent Salary Disadvantage	Sample Size[a]
Anatomy	10.2	$17,340	.034	13	130
Biochemistry	9.5	17,370	.065	24	542
Biophysics	5.5	18,930	.054	16	57
Botany	9.4	15,210	.033	10	261
Ecology	3.8	15,940	.039	8	202
Entomology	2.1	16,200	.040	19	145
Genetics	9.6	16,240	.054	18	235
Microbiology	13.6	16,670	.065	14	278
Nutrition	16.3	17,800	.034	6	177
Pharmacology	9.0	16,800	.077	14	129
Physiology	9.4	17,040	.050	15	487
Zoology	8.4	15,770	.039	16	282

[a] Women were sampled at a rate of 100 percent, men at a rate of 33 percent.

women, and have salaries which are higher (and steeper with experience), while biology and sociology have more women and have salaries which are lower (and flatter with experience). Given the large number of women in our biology sample, we sought to find out whether *within* biology women are concentrated in several low paying specialties with the apparent male-female differential being defined by the relative weights of men and women in these different "suboccupations." In Table 9 we have data on average gross professional earnings, percent wage disadvantage for females,[30] percent female, and steepness of the estimated profile at $XPO=0$.

In contrast to our disaggregation by school type there is no apparent relation between percent female and professional income or the steepness of the earnings profile ($\partial ln\ SAL/\partial XPO|_{XPO=0}$). In fact, the percent female and gross professional earnings are positively related ($r=.210$) and the percentage wage disadvantage and percent female are negatively related ($r=-.30$).[31] In conclusion while disaggregating by research orientation of the school relates to percent female, a disaggregation by suboccupation reveals no overrepresentation of women in the lower paying areas.

[30] While our interactive form worked for most subsamples, the small sample sizes motivated the introduction of *FEM* as a simple shift variable.

[31] Given the small sample size (12) these correlations are not, however, statistically significant.

E. *Earnings Differentials by Type of Employer*

Is the earnings disadvantage of women greater in certain types of employment and is the differential more experience dependent in certain sectors? The latter question cannot be answered because of the rather small number of Ph.D.s outside of academics, which limits fitting the interactive form used for Table 2. In Table 10 we have summarized the gross professional earnings differential by sector, and two features are of interest. First, the differentials in gross earnings are much larger than in salary. This can be interpreted as reflecting the smaller desired labor supply of women, but it is also consistent with discrimination against women in consulting opportunities. A second notable feature is the smaller differential in the government sector. It could be that governmental agencies, having greatest public visibility among the sectors, are unwilling to allow large differentials between men and women even if the women were less productive because of obsolescence of skills during the years of partial withdrawal from the labor force. An alternative interpretation is that government agencies resolutely avoid discrimination while the other sectors do not.

TABLE 10—ESTIMATED COEFFICIENTS ON *FEM* IN REGRESSIONS USING GROSS EARNINGS AS DEPENDENT VARIABLE FOR DIFFERENT DISCIPLINES BY TYPE OF EMPLOYER, 1970

	Academics	Government	Business
Economics	−.228	−.055	−.222
	(.022)	(.049)	(.143)
Sociology	−.174	−.076	[a]
	(.022)	(.090)	
Mathematics	−.251	−.075	−.302
	(.017)	(.080)	(.080)
Biology	−.213	−.195	−.245
	(.009)	(.046)	(.028)
Physics	−.217	−.156	−.217
	(.024)	(.040)	(.048)

[a] Insufficient number of observations.

IV. Conclusion

In this paper we have presented evidence on various aspects of the status of women faculty relative to men. The most important finding is that the academic salaries of females start out at not much less than those of males (4 to 11 percent less in the six disciplines in our sample) and then decline to result in a fairly substantial differential after a number of years of potential experience (13 to 23 percent at 15 years after the completion of the doctorate). In addition, we show that women academics are more likely than men to be employed by institutions which emphasize teaching rather than research and graduate training. These results would be expected to prevail even in the absence of direct labor market discrimination against women, for many women interrupt their careers through labor force withdrawal and/or part-time employment (or, perhaps as important, expect to do so *ex ante*) during the child-rearing years.

That the salaries of female academics relative to those of male academics decline with years since receipt of the Ph.D. up to about age forty-five is a matter of fact; what prudent persons can disagree on is just what this evidence means. The two principal alternative explanations are that the sex differences in salaries reflect a) differences in acquired skill and productivity between men and women, and b) direct labor market discrimination against women by male-dominated university faculties and administrations.[32] It is, of course, also

[32] It could also be argued that in a research oriented academic setting part of the male-female differential can be attributed to *positive* discrimination in favor of less competent males. Consider three professors: a male and a female of equal research ability and a male of less ability. (All are good teachers.) If the female chooses to do only teaching because of family commitments, she will receive a lower wage if the department values research. Though both males attempt to do research as well as teach, only one is in fact successful at research. The department will have to pay the going wage to keep the successful researcher, but as a utility maximizer the department chairman might be reluctant to

TABLE 11—ESTIMATED FEMALE SALARY DISADVANTAGE BY YEARS SINCE RECEIPT OF DOCTORATE AND FRACTIONS OF DISADVANTAGE ATTRIBUTABLE TO DISCRIMINATION AND TO HUMAN CAPITAL DIFFERENCES

	Years Since Receipt of Doctorate						
	0	5	10	15	20	25	30
Average Percentage Salary Disadvantage	6.9	11.8	15.5	18.3	20.0	21.0	21.0
Fraction Attributable to:							
Discrimination	1.00	.58	.45	.38	.35	.33	.33
Human Capital Differences	0	.42	.55	.62	.65	.67	.67

possible that the salary disadvantage of women is attributable to a combination of both of these factors.

From our study can one estimate the fraction of the observed salary disadvantage of women academics due to discrimination versus human capital differences? If one assumes that upon completion of the doctorate women and men have equal stocks of human capital and choose to rent them out in equal proportion, then the salary differential at the point of completion of the doctorate[33] may be interpreted as the "discrimination coefficient," for at this time there has presumably been little on-the-job training or obsolescence.

As is shown in Table 11, the simple average salary disadvantage of women in our sample of disciplines is 6.9 percent at $XPO=0$. This implies that the fraction of the salary disadvantage which is attributable to discrimination declines from 1.00 to .33 near retirement. Over a thirty-five-year work life, roughly two-fifths of the wage disadvantage is attributable to discrimination, three-fifths to human capital differences.[34]

If one accepts the conclusion that over half of the academic salary differential by sex can be explained by the market's reaction to voluntary choices by females regarding on-the-job training, then the implementation of antidiscrimination policies can be reconsidered. First, differentials arising through part-time experience or nonparticipation on the part of women need not reflect discrimination. Second, intrafamily choices of family planning and housework time which are much less subject to direct policy action are likely to be more important in changing the male-female salary structure in the long run. Third, in schools which have a strong research orientation those women faculty who do seek out research opportunities for periods of child care and part-time employment should find a smaller erosion of their marketable skills. Hence, when they return to full-time employment these women

tell the unsuccessful researcher that his published output is unimportant (by paying him a salary identical to that of the woman). Hence the chairman will overpay him, and this of course would result in the same empirical observation as discrimination against women. This tendency would be expected to prevail more strongly in a not-for-profit organization such as an academic department than in an organization in a competitive market.

[33] These assumptions are likely to be erroneous in two respects, but the errors work in opposite directions. That is, women would have accumulated less capital in graduate school if they anticipate family obligations, but these expected obligations should induce them to rent out a higher fraction of their capital when they first enter the labor force.

[34] This disaggregation of the differential is comparable to that found in a study of male-female pay differentials in a professional organization. Over the different samples studied, about 60 percent of the differential could be attributed to the different acquired skill characteristics of men and women. See Malkiel and Malkiel, p. 701.

should have salaries which more closely match those of their male colleagues.

REFERENCES

H. Astin, *The Woman Doctorate in America,* New York 1969.

G. Becker, *The Economics of Human Capital,* New York 1964.

Y. Ben-Porath, "The Production of Human Capital and the Life Cycle of Earnings," *J. Polit. Econ.,* Aug. 1967, *75,* 352–64.

B. Bergmann, "The Effect on White Incomes of Discrimination in Employment," *J. Polit. Econ.,* Mar./Apr. 1971, *79,* 294–313.

A. M. Cartter, *An Assessment of Quality in Graduate Education,* Washington 1966.

L. R. Harmon, *High School Ability Patterns, A Backward Look from the Doctorate,* Washington 1965.

R. W. Hodge and P. Hodge, "Occupational Assimilation as a Competitive Process," *Amer. J. Soc.,* Nov. 1965, *71,* 249–364.

G. E. Johnson and F. P. Stafford, "Social Returns to Quantity and Quality of Schooling," *J. Hum. Resources,* Spring 1973, *8,* 139–55.

——— and ———, "Lifetime Earnings in a Professional Labor Market: Academic Economists," *J. Polit. Econ.,* forthcoming 1974.

D. A. Katz, "Faculty Salaries, Promotions, and Productivity in a Large University," *Amer. Econ. Rev.,* June 1973, *63,* 469–77.

E. M. Landes, "Male-Female Differences in Wages and Employment: A Specific Human Capital Model," unpublished doctoral dissertation, Columbia Univ. 1974.

B. Malkiel and J. Malkiel, "Male-Female Pay Differentials in Professional Employment," *Amer. Econ. Rev.,* Sept. 1973, *63,* 693–705.

J. Mincer, *Schooling, Experience and Earnings,* forthcoming New York 1974.

——— and S. Polachek, "Family Investments in Human Capital: Earnings of Women," *J. Polit. Econ.,* forthcoming 1974.

S. Rosen, "Learning and Experience in the Labor Market," *J. Hum. Resources,* Summer 1972, *7,* 326–42.

L. C. Solmon, "Women in Graduate Education: Clues and Puzzles Regarding Institutional Discrimination," *Research in Higher Education,* Dec. 1973, *1,* 299–332.

P. Stephan, "Life Cycle Training Paths Subsequent to Formal Education," unpublished doctoral dissertation, Univ. Michigan 1971.

S. Warkov, A. Berger, and B. Frisbie, *Graduate Student Finances, 1963,* Chicago 1965.

F. Welch, "Black-White Differences in Returns to Schooling," *Amer. Econ. Rev.,* Dec. 1973, *63,* 893–907.

Michigan Education Association, *Michigan State University Salary Analysis, 1972–73,* Lansing 1972.

National Academy of Sciences, Research Division, *Careers of Ph.D.'s: Academic Versus Nonacademic,* Washington 1968.

U.S. Bureau of the Census, *Census of Population in 1970: Occupational Characteristics,* Subject Report PC(2)-7A, Washington 1971.

The Expected Impact of the Wage-Price Freeze on Relative Shares

By Sol S. Shalit and Uri Ben-Zion*

This paper attempts to shed some empirical light upon the impact of the wage-price freeze of August 15, 1971 on the relative shares of business and labor. While a complete evaluation of the freeze's impact would have to await the availability of extensive data, our indirect approach—an analysis of market expectations—has the advantage of yielding immediate results which are perhaps suggestive of the real impact. Based on analysis of the postfreeze behavior of stock market prices, our empirical findings are that the market expected the wage-price freeze to improve the relative share of corporate profits at the expense of wages.

We begin with a general discussion of the subject, and proceed with the theoretical relationship between expected earnings, stock prices, and the freeze. We then formulate the empirical model, state the operational hypothesis, and describe the data. Finally, we present the statistical findings and some conclusions.

I. Effects of the Wage-Price Freeze

The imposition of the wage-price freeze has evoked from its very beginning a bitter controversy between the Administration and organized labor concerning the freeze's overall effectiveness and its likely impact on relative shares.

Clearly, if postfreeze supply and demand conditions remained unchanged, a freeze would be effective but redundant; prices and wages would not have been any different in its absence. However, if the underlying conditions do change in the postfreeze period, there would be a priori grounds to suspect that a wage freeze might be less costly to police and enforce than a price freeze: firms can evade the regulation, at least partially, by cancelling discounts, redesigning the product, etc.[1] In contrast, labor contracts, particularly those of unionized workers, are often negotiated under conditions of high public visibility in which there exists little incentive for firms to pay higher wages than the legal frozen ones. Indeed, it has been labor's contention that the freeze would do more to hold down wages than prices, thereby adversely affecting their relative share.

Casual observation from the early days of the freeze suggests that the public apparently did not regard the new incomes policy as being neutral between wages and profits:

1) The first postfreeze day witnessed a record price gain in the stock market indicating widespread anticipations of a favorable impact on business.[2]

* University of Wisconsin-Milwaukee and Technion—Israel Institute of Technology, and University of Minnesota, respectively. We have benefitted from helpful comments by Lawrence Fisher, Charles Freedman, John Hause, H. Gregg Lewis, Christopher Sims, N. J. Simler, George J. Stigler, and Neil Wallace. We want to thank Robert Mullen for computer programming. Needless to say, we alone are responsible for remaining errors.

[1] For the experience with wage-price controls during World War II, see Colin Campbell; for the current experience, see the *Wall Street Journal*, Aug. 3, 1972, p. 1.

[2] The overall rise of the market in itself cannot obviously identify the source of these expectations. Our empirical analysis, however, avoids this difficulty by operating on cross-sectional changes in individual stock prices, rather than on the determinants of the overall market rise.

2) The enthusiastic endorsement of the freeze in business circles as opposed to its utter condemnation by labor suggests that the affected parties anticipated a differential impact on profits and wages.[3]

3) The existence at that time of many proposals for the control of corporate profits implies a tacit admission that price controls would not be entirely effective.

If labor's claim is valid, we would expect to find a higher increase in postfreeze profits of labor-intensive firms than in similar, less labor-intensive ones. Since the lack of appropriate wage, price, and income data makes a direct short-run test impossible at this time, we have tried an indirect approach; rather than analyze relative shares directly, we have focussed on the market's expectations concerning the effect of the freeze. Economic theory suggests that earnings' expectations are immediately capitalized in stock prices. Thus, we tested the hypothesis that postfreeze returns to investors in the stock market were systematically and positively related to firms' labor intensities.[4]

II. The Theoretical Relationship Between Expected Earnings, Stock Prices, and the Freeze

Economic theory postulates that in a world of certainty the relationship between assets' prices and their future earnings is given by:

$$(1) \qquad P_{it} = \overline{E}_i/\rho_{it}$$

where P_{it} is the price of stock i at time t, $\overline{E}_i$ is the average long-run earnings per share per unit of time, and ρ_{it} is the firm's cost of capital per unit of time at t. Accordingly, for a short period of time Δt, the major determinants of stock price changes (cum dividend) ΔP_{it}, are changes in long-run earnings $\Delta\overline{E}_i$, and changes in the cost of capital $\Delta\rho_{it}$. Taking first relative differences in (1) we get:

$$(2) \qquad R_{it} = \Delta P_{it}/P_{it} = \Delta\overline{E}_i/\overline{E}_i - \Delta\rho_{it}/\rho_{it}$$

where R_{it} represents the return on stock i during the time period Δt as measured by its relative price change during this time period.

In a world of uncertainty, the condition for equilibrium in the stock market suggests that:

$$(3) \qquad \rho_{it} = R_{Ft} + \lambda_t X_{it} = R_{Ft}(1 + \delta_{it} X_i)$$

where:

R_{Ft} = the risk-free capitalization rate,
λ_t = the market risk premium per unit of risk,
X_{it} = an index measure of the relative riskiness of asset i,
$\delta_{it} = \lambda_t/R_{Ft}$ is the relative risk premium expressed in terms of the riskless rate.

Taking the logarithms and the differentials of both sides, equation (3) can be approximated to yield:

$$(4) \qquad \Delta\rho_{it}/\rho_{it} = \Delta R_{Ft}/R_{Ft} + \Delta\delta_{it} X_{it} + \delta_{it}\Delta X_{it}$$

Substituting (4) into (2) and assuming ΔX_{it} to constitute an independently distributed random factor ϵ_{it}, we get:

$$(5) \qquad R_{it} = \Delta\overline{E}_i/\overline{E}_i - \Delta R_{Ft}/R_{Ft} - \Delta\delta_{it} X_{it} + \epsilon_{it}$$

The first term represents changes in investors' expectations of the firm's long-run earning power; the second term, changes in the riskless rate; the third term, the effect of changes in the market risk premium; and the last term, an additive random factor.

The discussion in Section I suggests that changes in expected postfreeze earnings are a function of the firm's labor in-

[3] See *The New York Times*, *Wall Street Journal*, and *Business Week*, the week of Aug. 16–22, 1971.

[4] The basic framework of our approach has been outlined in our 1973 paper in Vogt and Mickle.

tensity l_i. This can be approximated by the relationship:

$$(6) \qquad \Delta\overline{E}_i/\overline{E}_i = \beta_0 + \beta_1 l_{it} + w_{it}$$

where w_{it} is a random error with expectation $E(\epsilon, w)=0$.

Substituting (6) into (5), and recognizing that in a cross-section analysis of firm data the change in the riskless rate, $\Delta R_{Ft}/R_{Ft}$, is constant for all firms, we obtain a testable equation:

$$(7) \qquad R_{it} = \gamma_0 + \gamma_1 l_{it} + \gamma_2 X_{it} + u_{it}$$

where

$\gamma_0 = \beta_0 - \Delta R_{Ft}/R_{Ft}$ is a constant,
X_{it} = measure of corporate risk,
$\gamma_2 = \Delta\delta_{it}$ is the change in the market risk premium which at time t constitutes a common factor affecting stock prices and returns proportional to their risk levels,
$u_{it} = w_{it} + \epsilon_{it}$ is a random error.

In economic terms, equation (7) presents a postfreeze relationship between changes in the firm's stock prices R_i and its labor intensity l_i, holding risk X_i constant. Thus, (7) is a direct implication of (2) in a world of uncertainty. Therefore, a positive sign of the labor-intensity coefficient $\gamma_1 > 0$ will be consistent with the hypothesis that the freeze, as judged by market expectations, had the effect of increasing profits at the expense of wages.[5]

The sign of the risk coefficient γ_2 is expected to be positive ($\gamma_2 > 0$) for a rising market and negative ($\gamma_2 < 0$) for a falling one.

III. The Empirical Model and the Data

In order to render equation (7) operationally suitable for empirical testing, we have included the size of the firm as an additional control variable. In general, one would expect large firms to bear the brunt of the controls because smaller firms are more costly to police and regulate. Furthermore, since firms with annual sales in excess of \$100 million were singled out by the regulator for special controls, we felt it desirable to analyze the independent influence of the firm's absolute size.[6] Thus, equation (7) becomes:

$$(8) \quad R_{it} = \alpha_0 + \alpha_1 L_{it} + \alpha_2 X_{it} + \alpha_3 S_{it} + v_{it}$$

where:

R_{it} = the return to investors in stock i during time period Δt,
L_{it} = the firm's labor intensity,
X_{it} = the firm's risk,
S_{it} = the firm's absolute size,
v_{it} = a stochastic error term.

Our basic sample for estimating the parameters of equation (8) was the 1,000 largest industrial corporations in the United States according to *Fortvne's* "Double 500 Directory" of 1970. The dependent variable R_{it} ("return"), defined as the short-run rate of change in stock prices, was measured by the logarithm of stock-price ratios during various time intervals for the postfreeze *RF* and the prefreeze *RB* periods. These time intervals, spanning a ten-month postfreeze period and a six-month prefreeze period, were arbitrarily chosen; subperiods of a short duration were selected close to the freeze.

[5] Since our testable equation (8) estimates parameters relating to price changes rather than price levels, the possibility of omitted variables cannot be serious. To the extent that a variable, say the expected rate of growth of the firm's earnings, G, is relevant for the valuation equation (1), only its change, $\Delta G/G$, would be relevant for equation (8). We assume for simplicity that during Δt, the short period of time under investigation, $\Delta G/G$ was either constant or that its variations were captured by the residual term u_{it}. Our results could be biased only if we excluded relevant variables whose values were changed specifically by the imposition of the freeze in a way which systematically correlates with variables included in our regressions, labor-intensity in particular. We can conceive of no such excluded variables.

[6] About one-quarter of our basic sample consists of firms below \$100 million annual sales.

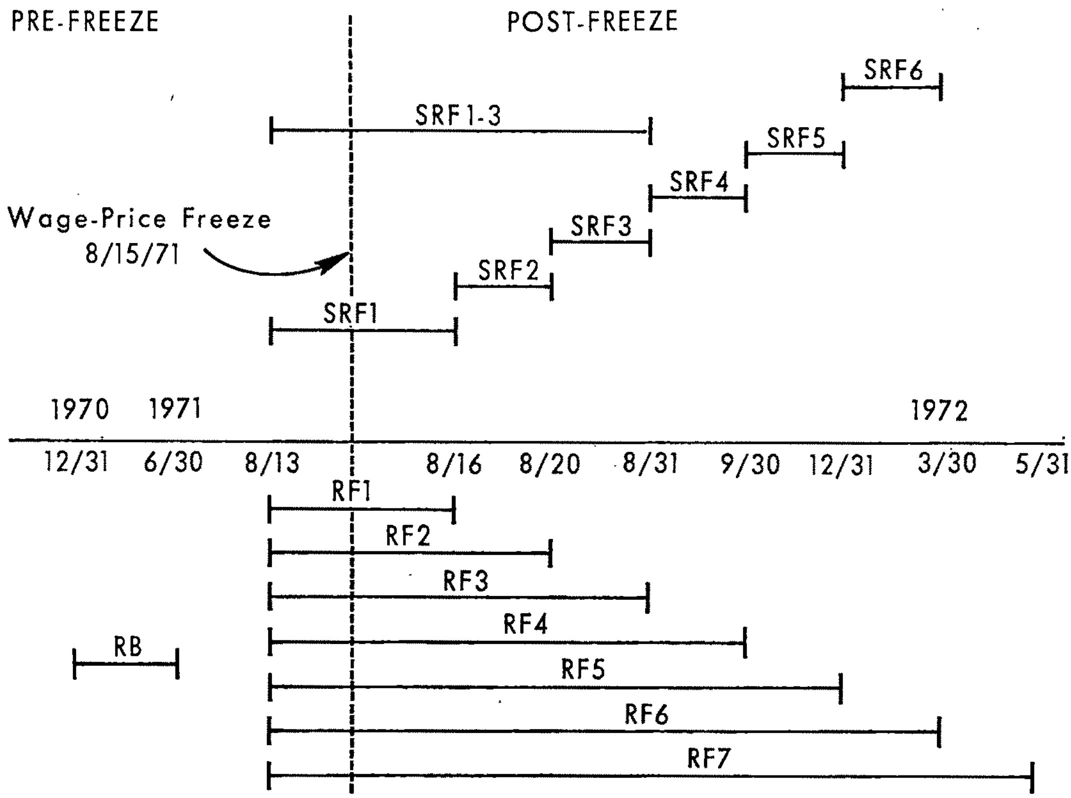


FIGURE 1. TIME INTERVALS FOR PREFREEZE AND POSTFREEZE RETURNS
(Time Intervals not drawn to Scale)

As a measure of the firm's labor intensity L_i, it would have been desirable to use the ratio of the firm's wage bill to its total cost. However, such data are not readily available and consequently we used the 1970 ratio of employees/sales, N/S.[7]

The risk variable X_i has been approximated by three alternative measures. The first was the widely used Standard and Poor's Earnings and Dividend Ranking, translated to numerical values ranging from 1 to 7 and based on Standard and Poor's *Stock Guide*. The second measure was *beta*, the regression coefficient of the individual stock's return over the market's return, a measure widely used in the academic literature of finance. The third measure used observations on stock returns during previous periods as proxies for risk in subsequent ones, assuming that risky stocks rise more in a rising market and fall more in a falling one. For the sake of brevity, only the results of earnings dividends ranking *EDRISK* have been reported below. The other risk measures gave substantially the same results, which have been reported in greater detail elsewhere.[8]

Two alternative measures were used to approximate the firm's absolute size S, the

[7] We have also tried the following ratios as proxy variables: employees/sales 1970, employees/sales 1971, average employees/sales 1970–71, employees/assets 1970, and employees/assets 1971. All these ratios have given substantially the same empirical results. Taking the ratio of employees/sales as a proxy for wage bill/sales is tantamount to assuming that the wage rate is constant across firms and industries. We have used the proxy in the absence of detailed wage data by firms.

[8] See our 1972 paper.

firm's 1970 sales, and its 1970 assets. Since both measures gave almost the same results, we have reported below the 1970 sales as a measure of size *SIZE*.

The main thrust of the paper is the independent effect of labor intensity on stock returns during seven postfreeze periods ($RF_1, \ldots, RF_7$). We also found it interesting to test whether labor intensity had a significant effect on stock returns before the wage-price freeze. For comparison, therefore, we have estimated the parameter of the model for a six-month prefreeze period (Dec. 1970–June 1971) during which the market as a whole rose (*RB*).[9] Finally, the unexpected imposition of the wage-price freeze offers a rare opportunity to study some dynamic aspects of stock market behavior. We deemed it valuable to analyze the speed with which the market was able to digest and act upon new information, and the length of time it took to complete the adjustments towards a new equilibrium. With this in mind, we have calculated additional postfreeze returns for intermediate short periods ($SRF1, \ldots, SRF6$).

Figure 1 illustrates the various time intervals covered by all returns: prefreeze, postfreeze, "short," and "long." Detailed definitions of all the variables and their sources can be found in the Appendix, which also provides the level of Standard and Poor's 425 Industrials for the corresponding time periods.

IV. Empirical Results

Equation (8) has been estimated by a cross-section set of *OLS* regressions; both linear and logarithmic forms gave substantially the same results. Table 1 presents the statistical results in *log* form for the seven postfreeze periods ($RF1, \ldots, RF7$), and the prefreeze period (*RB*).

The most remarkable feature of Table 1 is the positive and highly significant effect of labor intensity N/S on postfreeze stock returns, holding constant the firm's risk and size. The second most important feature is the persistence of this effect throughout the entire period under review, extending from the first postfreeze day (Aug. 16, 1971) to almost ten months later (May 31, 1972).[10] Finally, the sign and significance of the labor intensity regression coefficients were found to be invariant with the use of alternative measures of risk.[11]

The empirical results in Table 1 are consistent with the hypothesis that the market expected the wage-price freeze to increase profits of labor-intensive firms significantly when compared with less labor-intensive firms of similar size and risk. This conclusion is reinforced by the results reported in the last column of Table 1, which gives the estimated parameters of equation (8) for the prefreeze period and in which no significant relationship could be found between returns and labor intensity.

The rather low percentage of "explained" variation (R^2) was not surprising since changes in stock prices are subject to other random effects. Low levels of R^2 are quite common in studies which make use of cross-sectional changes in individual stock prices. However, rather than explain stock market returns, our objective here is to test a hypothesis. On the basis of the evidence reported here, we can reject the hypothesis that the market expected the freeze to have no effect in increasing the

[9] Similar empirical results were obtained for a prefreeze period during which the market as a whole declined; for details see our 1972 paper.

[10] It should be emphasized that the cutoff date (May 31, 1972) was entirely arbitrary and should not be interpreted to mean that the observed effect did not last more than ten months. We simply did not extend our study beyond that date.

[11] For full details on regressions results using other measures of risk, see our 1972 paper.

TABLE 1—POSTFREEZE AND PREFREEZE RETURNS ON LABOR INTENSITY, RISK, AND SIZE
(t-values in parentheses)

	$RF1$	$RF2$	$RF3$	$RF4$	$RF5$	$RF6$	$RF7$	RB
Constant	.0394	.1077	.1066	.1192	.2723	.4544	.3464	.1097
	(2.160)	(4.624)	(2.910)	(2.174)	(3.840)	(4.302)	(3.089)	(1.141)
$\log (N/S)$	.0223	.0345	.0385	.0185	.0444	.0631	.0616	−.0104
	(6.137)	(7.465)	(5.288)	(1.702)	(3.155)	(3.008)	(2.765)	(.5429)
$\log (EDRISK)$	.0337	.0231	.0291	−.0012	−.0382	−.0093	−.1210	.0378
	(7.115)	(3.822)	(3.059)	(.084)	(2.076)	(.3414)	(.4164)	(1.513)
$\log (SIZE)$	.0080	.0040	.0073	−.0051	−.0079	−.0231	−.0135	−.0141
	(4.332)	(1.844)	(1.987)	(.9566)	(1.151)	(2.231)	(1.235)	(1.499)
F	30.9	24.9	13.0	1.55	4.68	5.80	4.1	2.99
Degrees of Freedom	719	719	719	719	719	719	719	719
R^2	.115	.095	.052	.006	.019	.024	.017	.012

relative share of profits at the expense of wages.

Concerning the size variable, it is not a priori clear what its sign should be. On the one hand, large firms may possess the advantage of a legal and professional staff to fend off legal and administrative interference. On the other hand, large corporations also constitute easier targets for the regulator due to the low cost of control. Indeed, the postfreeze results may suggest that as information began to accumulate, the market revised its expectations to be consistent with the special controls imposed on large firms. This interpretation gains some support from the results of another study (see our 1973 paper) in which we ran the regressions of Table 1 for a split sample with separate regressions for firms with annual sales above and below $100 million sales. The results for the size coefficients show that while the larger corporations continued to display reduced significance and reversal of signs, the smaller companies exhibited a consistent pattern of negative signs and low statistical significance throughout the entire postfreeze period under study. However, the independent effect of size is by no means clear, and more work is called for.

V. A Strong Test of the Model, Using Industry Dummies

One might argue that our empirical results need not necessarily be attributed to the wage-price freeze per se, but may, in part, be the consequence of the new economic policy which was enunciated jointly with the freeze, and which may have effected specific industries differently. We have therefore used industry dummy variables to take into account this industry affect, and to serve as proxies for other variables not directly measured, such as industry concentration, degree of unionization, etc.

However, the use of industry dummies poses some difficulties. An empirical finding of industry effects could be attributed either to a governmental policy specific to the industry or to its labor intensity; regression results alone cannot differentiate between the underlying causes. Furthermore, if industry classification explains a major portion of variations in labor intensity among firms, then the use of industry dummies should significantly reduce the observed effect of the labor intensity specific to the firm. In an extreme case where all firms in an industry have the same underlying labor intensity,

TABLE 2—POSTFREEZE AND PREFREEZE RETURNS WITH INDUSTRY DUMMY VARIABLES
(t-values in parentheses)

	*RF*1	*RF*2	*RF*3	*RF*4	*RF*5	*RF*6	*RF*7	*RB*
Constant	.0076	.0656	.0890	.0351	.2200	.4273	.2805	−.0159
	(.3494)	(2.379)	(2.001)	(.5206)	(2.540)	(3.315)	(2.084)	(.1370)
log (*N/S*)	.0111	.0198	.0287	−.0088	.0501	.0566	.0346	−.0396
	(2.392)	(3.377)	(3.034)	(.6155)	(2.722)	(2.067)	(1.210)	(1.605)
log (*EDRISK*)	.0262	.0122	.0152	−.0169	−.0311	−.0143	.0035	.0443
	(5.172)	(1.916)	(1.487)	(1.086)	(1.553)	(.4819)	(.1121)	(1.649)
log (*SIZE*)	.0072	.0029	.0038	−.0082	−.0069	−.0275	−.0179	−.0168
	(3.924)	(1.248)	(1.022)	(1.455)	(.9468)	(2.531)	(1.577)	(1.716)
F	6.75	6.83	3.60	1.3	2.1	2.3	3.17	2.6
Degrees of Freedom	719	719	719	719	719	719	719	719
R^2	.189	.191	.11	.04	.068	.074	.098	.083

the differences among them would reflect measurement errors or transitory differences during an adjustment process. In either case, we would not expect to find any significant effect of the firm's labor intensity, although a positive sign would be perfectly valid on theoretical grounds. Thus, the reestimation of equation (8) with industry dummies constitutes a strong test of our model.

We have classified the firms in our sample into twenty-one two-digit industries by using a three-digit industry classification.[12] We then reestimated equation (8) as follows:

$$R_{it} = \alpha_0 + \alpha_1 L_i + \alpha_2 X_i + \alpha_3 S_i + \sum_{j=1}^{21} \mu_{it} I_{ij} + v_{it} \tag{9}$$

where $I_j = 1$ if firm i belonged to industry j, and $I_j = 0$, otherwise.

The estimated parameters of the three main variables in (9) for seven postfreeze periods (*RF*1, . . . , *RF*7) and the prefreeze period (*RB*) are given in Table 2, and can be compared with those reported in Table 1.

The reestimated regressions show significant industry effects and a substantial increase in the level of R^2. Some industries display significant effects throughout (for example, machinery and transportation equipment); others have significant effects only during a few periods (for example, primary metals, textiles); and still others display no significant effect during any of the periods under study. As expected, the coefficient of labor intensity declined when industry dummies were used, yet the coefficient retained its correct sign in six out of the seven postfreeze regressions and was significant in five. The negative coefficient during the fourth period (*RF*4) was not significant and the level of explanation of the entire regression was somewhat lower; we have no clear explanation for this.[13] The overall results of the strong test, however, are consistent with our model.

Finally, we attempted to test our contention that the industry dummies may have partially captured differences in the industry's labor intensity. We have analyzed the relationship between the coeffi-

[12] See "Directory of Companies Filing Annual Reports with the Securities and Exchange Commission 1970."

[13] One possible explanation may be that in the middle of Phase 1 of the freeze (September 30) some news caused investors to expect an end to, or a substantial relaxation of, the freeze. We have no strong evidence for this assertion other than to point out that during the entire period under study, the only significant reversal of the Standard and Poor's Industrials Index occurred in this fourth subperiod, which ended on Sept. 30, 1971 (see Standard and Poor's Industrials in the Appendix).

TABLE 3—REGRESSION COEFFICIENTS OF INDUSTRY DUMMIES ON $LOG\ (N/S)_j$
(t-values in parentheses)

	$\hat{\mu}_{RF1}$	$\hat{\mu}_{RF2}$	$\hat{\mu}_{RF3}$	$\hat{\mu}_{RF4}$	$\hat{\mu}_{RF5}$	$\hat{\mu}_{RF6}$	$\hat{\mu}_{RF7}$	$\hat{\mu}_{RB}$
Constant	.0018	.0084	.0175	.0256	.0699	.0352	−.0194	.0017
	(.5314)	(1.710)	(3.021)	(3.907)	(5.411)	(1.490)	(.7543)	(.0908)
log (N/S)	.0179	.0264	.0169	.0469	−.0127	−.0132	.0444	.0461
	(1.980)	(2.126)	(1.100)	(2.704)	(.3704)	(.2116)	(.6516)	(.9387)
F	3.9	4.5	1.2	7.3	.137	.044	.42	.86
Degrees of Freedom	20	20	20	20	20	20	20	20
R^2	.17	.19	.06	.28	.007	.002	.021	.043

cients of the industry dummies and the mean labor intensity in the industry. For each of the seven postfreeze periods and the prefreeze period, we ran the following regressions:

$$\hat{\mu}_{jt} = \theta_0 + \theta_{it}(\overline{log\ N/S})_j + u_{jt} \tag{10}$$

where $\hat{\mu}_{jt}$ is the estimated regression coefficient of the industry dummy in equation (9), and $(\overline{logN/S})_j$ is the average *log* of labor intensity in the industry ($j=1, \ldots, 21$).

The regression coefficients of equation (10) are reported in Table 3. The statistical results indicate that throughout the first four postfreeze periods there was indeed a significantly positive relationship between the industry-dummy coefficients and the industry's labor intensity. However, no such relationship was found for the last three postfreeze periods and for the prefreeze period. Indeed, one would expect that as we move further away from the inception of the freeze, industry factors would depend more upon news specific to the industry, and less upon its average labor intensity. The coefficients of the industry-dummy variables (not reported here due to space limitations) behave accordingly: they are far more significant during the first four periods than they are during the last three. To summarize, the additional evidence gathered from what we termed the "strong test" adds some support to the model and reaffirms our hypothesis.

VI. A Note on Stock Market Efficiency

In order to get the full benefit which the wage-price freeze offers in analyzing the dynamics of the stock market, a daily, perhaps even an hourly, analysis of stock market returns would have been warranted, at least for the immediate postfreeze period. Table 4, which covers only arbitrary and varying intervals of one day to three months, nevertheless provides some insight into the market's adjustment to new information.

The most obvious point in Table 4 is that the stock market does not take an attitude of "wait and see"; it reacts without delay and in the theoretically expected direction. Thus, the biggest adjustment toward a new equilibrium took place on the very first day after the freeze ($SRF1$) when high labor-intensity stocks made their largest price gain. The results also indicate that it took more than a day, but less than a week, for the market to complete its adjustment. Labor intensity still had a significantly positive effect on stock returns during $SRF2$ (Aug. 17–20, 1971), although its effect was weaker when compared with the one which occurred on the first day (Aug. 16, 1971). In practical terms this means that following the end of the first postfreeze trading day, there was at least one day, but no more than a few days, during which an opportunity still existed to make a profit by investing in

TABLE 4—POSTFREEZE "SHORT" RETURNS ON LABOR-INTENSITY, RISK, AND SIZE

(t-values in parentheses)

	*SRF*1	*SRF*2	*SRF*3	*SRF*1–3	*SRF*4	*SRF*5	*SRF*6
Constant	.0394	.0661	−.0054	.0607	.0228	.1450	.2007
	(2.160)	(3.989)	(1.876)	(1.876)	(.4492)	(2.228)	(2.440)
log (N/S)	.0223	.0119	.0021	.0141	−.0157	.0208	.0229
	(6.137)	(3.656)	(.3715)	(2.225)	(1.581)	(1.640)	(1.425)
log (EDRISK)	.0337	−.0109	.0059	−.0050	−.0311	−.0385	.0286
	(7.115)	(2.520)	(.7635)	(.5958)	(2.340)	(2.263)	(1.328)
log (SIZE)	.0080	−.003	.0026	−.0069	−.0113	−.0042	−.0156
	(4.332)	(2.050)	(.9134)	(.2181)	(2.260)	(.6629)	(1.929)
F	30.9	7.07	.355	1.75	3.13	2.5	4.7
Degrees of Freedom	719	749	749	749	749	749	749
R^2	.115	.028	.001	.007	.012	.010	.019

high labor-intensity firms, despite an overall stable market. After the first trading week, however, no further opportunities existed to act profitably upon this information: there was no longer a significant relationship between labor intensity and stock returns in the period Aug. 20–31, 1971 (*SRF*3).[14]

In judging the extent to which our findings are consistent with the concept of an "efficient market," it is important to bear in mind that the observed market performance constitutes an underestimate of its actual efficiency in several respects:

First, the time period covered under *SRF*2 (Aug. 16–20, 1971) lumps the cumulative effect of four trading days which become an upper limit for the adjustment's duration. A finer breakdown of the data into daily short returns might conceivably show that the market had already completed the adjustment on the second postfreeze trading day (Aug. 17, 1971).

Second, it is unlikely that all the relevant information about the freeze was known to investors on the first day and that the subsequent market adjustments were a reaction to this information only. It is more likely that the dissemination of news was not an instantaneous process, that new information continued to reach the market during the week following the President's announcement, and that the market's adjustments were responses to new information.

Third, the observed duration for the market's adjustment process is a function of two components: the time it takes for the market to react and the time it takes for the stock exchanges to register that reaction. Since the exchanges' technical and administrative facilities are of a given capacity, not a frictionless theoretical construct, it is not surprising that the first postfreeze day, which witnessed a record-high volume, was characterized by substantial delays in the tape, suspension of trading in certain stocks, etc. Therefore, the observed performance during a substantial peak load strain is probably an underestimate of the market's efficiency

[14] The regression results for *SRF*4, *SRF*5, and *SRF*6 have been given merely for comparison with the long returns reported in Table 1. Their interpretation can be speculative at best. The further away we get from *SRF*1, the less can the results be attributed to the freeze and the more to bits of new information that continued to flow and affect the market. Thus, the negative effect during *SRF*4 could possibly be interpreted as a correction to an "overshooting" during *SRF*1–3, or alternatively, as a reaction to new information not necessarily connected with the freeze. In contrast, no such problems exist in the interpretation of the long returns (*RF*1, . . . , *RF*7) if the cumulative effect is strong. The decrease in the significance level of the employment-intensity coefficient from *RF*1 to *RF*7 can be attributed to a presumably increasing variance of the error term during the ten-month period over which these returns extend.

under normal conditions. However, our findings on market dynamics are not exhaustive, but rather suggestive of a need for more work in that direction.

VII. Summary and Conclusions

Our empirical results show that the market expected the wage-price freeze to improve the relative share of corporate profits at the expense of wages. Our results also show that these market expectations have not reversed themselves, but rather persisted throughout the ten-month postfreeze period under study. Postfreeze stock market returns of firms were found to be significantly and positively related to their labor intensities, holding size and risk constant. In contrast, no such relationship has been detected during the prefreeze period.

The stock market took its biggest stride towards a new equilibrium on the very first postfreeze day, while taking between a minimum of two and a maximum of four trading days to complete the adjustment.

Finally, as Hendrik Houthakker has already pointed out, economists are likely to debate the freeze's impact for years to come. Admittedly, our procedure does not constitute a *direct* test of the relative shares hypothesis, but of market expectations. However, according to economic theory these expectations do not form in a vacuum, but in the context of all available information; in a rational market these expectations are an unbiased prediction of the future. In this respect, our indirect test may not be irrelevant to the eventual analysis of the wage-price freeze policy and its impact on relative shares.

Appendix

Definition of Variables and Their Sources

Variable	Symbol	Definition[a]	*S&P* Industrial Average Index
	*RF*1	*log* [*P*(8-16-71)/*P*(8-13-71)]	108.92/105.46
	*RF*2	*log* [*P*(8-20-71)/*P*(8-13-71)]	108.49/105.46
	*RF*3	*log* [*P*(8-31-71)/*P*(8-13-71)]	109.40/105.46
Postfreeze returns RF_{it}	*RF*4	*log* [*P*(9-30-71)/*P*(8-13-71)]	108.77/105.46
	*RF*5	*log* [*P*(12-31-71)/*P*(8-13-71)]	112.72/105.46
	*RF*6	*log* [*P*(3-30-72)/*P*(8-13-71)]	119.26/105.46
	*RF*7	*log* [*P*(5-31-72)/*P*(8-13-71)]	122.49/105.46
	*SRF*1	*log* [*P*(8-16-71)/*P*(8-13-71)]	108.92/105.46
	*SRF*2	*log* [*P*(8-20-71)/*P*(8-16-71)]	108.49/108.92
	*SRF*3	*log* [*P*(8-31-71)/*P*(8-20-71)]	109.40/108.49
Postfreeze short returns SR_{it}	*SRF*4	*log* [*P*(9-30-71)/*P*(8-31-71)]	108.77/109.40
	*SRF*5	*log* [*P*(12-31-71)/*P*(9-30-71)]	112.72/108.77
	*SRF*6	*log* [*P*(3-30-72)/*P*(12-31-71)]	119.26/112.72
	*SRF*1–3	*log* [*P*(8-31-71)/*P*(8-16-71)]	109.40/108.92
Prefreeze returns RB_{it}	*RB*	*log* [*P*(6-30-71)/*P*(12-31-70)]	109.95/100.90
Employment-Intensity L_i	*N/S*	Employees 1970/Sales 1970	
Risk X_i	*EDRISK*	Earnings Dividend Ranking	
Size S_i	*SIZE*	Sales 1970	

[a] *P* represents stock prices on the designated dates, see Figure 1.

Source: Stock price data for calculating returns were derived from Standard and Poor's *Stock Guides* for the appropriate dates. Earnings and Dividend Ranking data were derived from the Aug. 1971 issue of the *Stock Guide*. Employment and sales figures were derived from *Fortune*, 1971.

REFERENCES

U. Ben-Zion and S. Shalit, "The Wage-Price Freeze: An Empirical Test of Short-Run Market Expectations," in G. Vogt and M. H. Mickle, eds., *Socio-Economic Systems and Principles*, Pittsburgh 1973.

C. D. Campbell, *Wage Price Controls in World War II, United States and Germany*, Washington 1971.

H. S. Houthakker, "Are Controls the Answer?," *Rev. Econ. Statist.*, Aug. 1972, *54*, 231.

S. Shalit and U. Ben-Zion, "The Impact of the Wage-Prize Freeze on Relative Shares: A Test of Short-Run Market Expectations," disc. paper no. 24, Univ. Minnesota, Dec. 1972.

Business Week, Aug. 16–22, 1971.

Fortune, "Double 500 Directory of 1,000 Largest U.S. Industrial Corporations," New York 1971, 1972.

New York Times, Aug. 16–22, 1971.

Standard and Poor's, Stock Guides, New York 1971, 1972.

U.S. Securities and Exchange Commission, "Directory of Companies Filing Annual Reports with the Securities and Exchange Commission 1970," Washington 1971.

Wall Street Journal, Aug. 16–22, 1971; "Firms Find New Benefits for Key Men," Aug. 3, 1972.

Pay and Performance in Major League Baseball

By Gerald W. Scully*

The 1972 baseball season opened with the first players' strike in baseball history. The issues were economic, focusing primarily on the player pension fund. In recent years, there has been growing discontent among players over the provisions of the uniform contract in organized baseball, an issue which threatened a strike and a lockout prior to the 1973 season. The principal issues remain minimum salaries, maximum salary cuts, player benefits, and the reserve clause. The players support the recent Curt Flood claim that the reserve clause is illegal and severely reduces player economic benefits. Economic analysis confirms the latter view (see Simon Rottenberg). The owners contend that the clause is necessary for equalization of playing strength among teams. Economic analysis disputes such a contention (see Mohamed El Hodiri and James Quirk). Further, owners insist that player salary demands are unrealistically high. Fans, caught in the middle, seem persuaded that the players are overpaid. Somehow, a $150,000 a year ball player claiming exploitation is anti-heroic.

* Professor of economics, Southern Methodist University. This paper evolved from my paper, "Discrimination: The Case of Baseball," which was prepared for the Brookings Conference on Government and the Sports Business, Dec. 6 and 7, 1971. I acknowledge that the participants at the conference and others, particularly, Josef Hadar, Roger Noll, James Quirk, Leonard Rapping, Sherwin Rosen, and the managing editor contributed to the ideas in this paper. In its early phases, this study was partially financed by PHS Research Grant No. 1-R01-HD05989-01 from the National Institute of Health. In its final stages, the study was partially financed by Institutional Grant No. 31-46-70-06 from the Manpower Administration, U.S. Department of Labor.

The purpose of this paper is to crudely measure the economic loss to the players due to the restrictions of the reserve clause. Relationships between player performance and salary will be estimated and the predicted salary compared to predicted player marginal revenue product. Aside from providing insights into the operation of a particular labor market, the approach differs from other studies of the wage determination process in the following ways: 1) the individual player is the unit of observation; 2) salary functions are estimated within the framework of a labor market characterized by extensive bargaining; 3) marginal revenue products of the factors of production are estimated explicitly; and 4) rates of monopsonistic exploitation are found by comparing salary and marginal revenue product over various performance levels and career lengths.

The discussion will unfold as follows. First, the institutional characteristics of the baseball players labor market are briefly outlined. Next, with this institutional framework in mind, a model of marginal revenue product and salary determination is formulated. Then, the results of the empirical investigation into marginal revenue product and salary determination are presented. Finally, the rate of monopsonistic exploitation is calculated and policy implications discussed.

I. The Organization of the Baseball Players Labor Market

The institutional framework governing the organization of the labor market in baseball has been discussed in depth previ-

ously (see Rottenberg). We will deal only briefly with the topic. Currently, a player is a *free agent* until he signs a one-year, renewable, *uniform contract* with a major league team. The renewable feature of the contract is an option granted exclusively to the owner and is known more widely as the *reserve clause.* Simply stated, the reserve clause restricts the player's freedom of negotiation to the owner of the contract. Short of a constraint restricting the owner in the percent that he can cut a player's salary from the previous season, the contract grants a license to the owner to dispose of the player's services in the manner he sees fit. The owner's options at the expiration of the contract are wide. He can renew, sell, or terminate the contract. If he renews the contract, the player may be relegated to the team roster, transferred to a minor league team on the "protected" list, or left eligible for the *player's draft.* In selling the contract to another team, the owner transfers a property right—the exclusive control over the player's services. If sold, the player must report to his new owner within seventy-two hours.

The player's options are more limited. Upon being offered a contract, the player can accept the terms and sign or attempt to negotiate an improvement in the terms in his favor. The player may resort to a number of ploys, including the "holdout" tactic, to improve the terms. But once the owner has made his "final" offer, the player must capitulate or withdraw from organized baseball.

The reserve clause plays an important role in the determination of player salaries. If the labor market in organized baseball were perfectly competitive, player salaries would be equated with player marginal revenue products (*MRP*). The reserve clause prohibits the transfer of player services to teams where their *MRP* is higher. The reserve clause restricts player bargaining to one owner. The restriction grants some monopsony power to the owner and the exercise of that power results in a divergence between *MRP* and salary. The marginal revenue product continues to be an essential factor in player salary determination, but under the reserve clause, players and owners share the player's *MRP*.

The player's marginal revenue product in baseball is the ability or performance that he contributes to the team and the effect of that performance on gate receipts. The effect of player performance on revenue may be direct or indirect. Ability contributes to team performance and victories raise gate receipts—and broadcast revenues; this is the substantial effect of the individual's performance. Additionally, it is possible that some players may attract fans over and above their individual contribution through the team. Both players and owners understand the relationship, and the player's ability is the key issue in bargaining. But aside from the actual *MRP* of the player, the relative strength of the bargainers will also affect salary. It is true that the reserve clause grants more economic power to the owner. However, management is constrained by considerations of player and team morale. Certain factors, such as the number of close substitute players, affect the outcome of the negotiation. But the personalities of the bargainers must contribute to relative bargaining strength. Bargaining always has in it an element of poker and, as in poker, the player is at least as important as the cards.

II. A Simple Model of Marginal Revenue Product and Salary Determination in Major League Baseball[1]

A model which purports to reveal simply the formal process of *MRP* and salary determination should incorporate the two

[1] For an elegant treatment of the theory, see El Hodiri and Quirk.

major features of the ball players labor market discussed above: 1) Gross baseball revenues are related to individual performance primarily through their effect on team standing; and 2) the reserve clause reduces player salaries below player marginal revenue products.

Teams are engaged in the production of a constant number of games of a certain level of quality. The quality of the games is assumed to be measured by the team's percent wins W, which is related to two general categories of inputs: 1) a vector of player skills A_i; and 2) a vector of other nonplayer inputs I_j, such as managers, coaches, capital, etc., and more nebulous inputs such as team spirit. Thus,

$$(1) \quad W = W(A_1, A_2, \ldots, A_n; I_1, I_2, \ldots, I_m)$$

Teams derive revenue essentially from two main sources: gate receipts and the sale of radio and television rights. Fans purchase tickets or watch televised games because they derive utility from seeing the home team win. We assert that both gate receipts and broadcast revenue are directly related to the team's percent wins and the population in the area and indirectly related to player performance. That is, fans attend or watch games to see the team win, not to see player skills per se. It is well known that attendance varies with team performance and size of the metropolitan area (see Roger Noll), and undoubtedly so do broadcast revenues (see I. Horowitz). Accordingly, we state

$$(2) \quad \begin{aligned} R &= p \cdot T[W(A_i, I_j), P_g] \\ &\quad + B[W(A_i, I_j), P_b] \\ &\quad i = 1, \ldots, n; \quad j = 1, \ldots, m \end{aligned}$$

where R = team revenue, p = ticket price, T = number of tickets sold, W = team performance, P_g = population potentially attracted to the ballpark, P_b = potential broadcast households, and B = broadcast revenues.

Team costs C will be determined by the level of skills of the players and by the level of other nonplayer inputs. Since teams face a monopsonistic labor market, player costs are endogenously related to the level of skill. Other factor markets are assumed competitive. Therefore, we define team costs as

$$(3) \quad \begin{aligned} C &= \sum_i A_i S_i(A_i) + \sum_j r_j I_j \\ &\quad i = 1, \ldots, n; \quad j = 1, \ldots, m \end{aligned}$$

where $S_i(A_i)$ are player supply functions and r_j nonplayer factor remunerations.

Team profits π are defined as

$$(4) \quad \pi = R - C$$

The first-order conditions for a maximum are obtained by differentiating with respect to A_i and I_j.

$$(5) \quad \begin{aligned} \frac{\partial \pi}{\partial A_i} &= p \frac{\partial T}{\partial W} \frac{\partial W}{\partial A_i} + \frac{\partial B}{\partial W} \frac{\partial W}{\partial A_i} \\ &\quad - A_i \frac{\partial S_i}{\partial A_i} - S_i, \\ &\quad i = 1, \ldots, n \end{aligned}$$

$$(6) \quad \begin{aligned} \frac{\partial \pi}{\partial I_j} &= p \frac{\partial T}{\partial W} \frac{\partial W}{\partial I_j} + \frac{\partial B}{\partial W} \frac{\partial W}{\partial I_j} - r_j \\ &\quad j = 1, \ldots, m \end{aligned}$$

The first-order conditions reveal that teams maximize profits by selecting a level of player skills and nonplayer inputs such that players receive a salary equal to their marginal revenue products less monopsony rents ($A_i S_i'$), and other factors are remunerated equal to their marginal revenue products. Owner accrued monopoly profits are also present in the model. The manifestation of monopoly profits may be considered by observing that

$$(7) \quad \frac{\partial \pi}{\partial P_g} = p \frac{\partial T}{\partial P_g} > 0$$

$$\frac{\partial \pi}{\partial P_b} = \frac{\partial B}{\partial P_b} > 0 \tag{8}$$

III. Player Marginal Revenue Products

It is possible to crudely estimate player marginal revenue products in major league baseball. While the marginal revenue product of a particular player cannot be determined readily, the effect of player performance on team winning and the effect of team winning on team revenue can be ascertained. Making reasonable assumptions about how a player's performance alters team performance permits approximations of the player's marginal revenue product. It would be better if the relationship between marginal revenue product and player performance could be estimated using the player as the unit of measurement rather than the team, as can be done in the salary regressions. Such a procedure is impossible for the hitters and very difficult for the pitchers. Hitters in the regular lineup play virtually every home and away game during the season. Hence, fans cannot discriminate through attendance on the basis of player appearance. Pitchers do not start every game, so it is possible to evaluate their marginal revenue products individually. But this requires laboriously analyzing attendance for every team for every game during the season.

Player marginal revenue products are estimated in a two-equation model.[2] Equation (9) in essence is the production function which relates team output, the win-loss percent, to a number of team inputs. Equation (10) is the team revenue function which relates team revenues to the team win-loss record and the principal market characteristics of the area in which the team plays. The justification for estimating these relationships separately is derived from the fact that fans attend games principally to see the home team win, not to witness hitting and pitching performance per se.[3] Equation (9) specifies a linear relationship between team percent wins *PCTWIN*, expressed conventionally as games won divided by total games times 1,000, and a number of team performance variables. On the basis of previous research into the determinates of player salaries, it was found that hitter and pitcher performance were most suitably measured by the slugging average *SA* and the strikeout-to-walk ratio *SW*, respectively.[4] Therefore, the team win-loss percent is hypothesized to be related to the team slugging average *TSA* and the team strikeout-to-walk ratio *TSW* which are assumed to measure the offensive and defensive contributions to team victories. Winning, however, may be determined by more than team hitting and pitching performance. A large number of games in the majors during the season are won by one run. In these instances, player and hitting performance per se will make less difference in the outcome. It is here that team play, hustle, the quality of managerial and on the field decision making, all of which is determined by team morale, will substantially determine which team wins a higher share of these one-run games. To adjust for these factors, two dummy variables are specified: *CONT* which is equal to one for pennant or divisional winners

[2] Anthony Pascal and Leonard Rapping came close to estimating player marginal products in baseball. They specified a two-equation model, one relating team performance to team wins, the other between team wins and attendance but did not interpret their results as estimates of player marginal products.

[3] One must concede, however, that fans are drawn to see superstar players in action. The point, however, is that team revenue will be correlated principally with team win-loss records and less readily with team performance statistics. This is verified by empirical evidence, that the zero-order correlation coefficient between team revenue and team percent wins is much higher than the correlation coefficients between team revenue and team hitting and pitching measures.

[4] See the author (1974).

and their closest competitors at the end of the season, if the competition was five or fewer games out; and *OUT* which is equal to one for teams which at the end of the season were twenty or more games out of placing. Variable *CONT* should capture quality of play above and beyond hitting and pitching performance, inspired by the potential of a pennant or divisional title. Variable *OUT* should capture the demoralization that surrounds habitual losers. Finally, the absolute quality of play is generally regarded as being higher in the National League. Thus, equivalent team hitting and pitching performance would be expected to yield a lower win-loss percent in the National League. To compensate for this factor, a dummy variable *NL*, equal to one if the team is in the National League, is specified.

The team percent win function *PCTWIN* was estimated with team data for 1968 and 1969.[5] The statistical results are:

$$(9)\quad PCTWIN_t = \underset{(.39)}{37.24} + \underset{(4.37)}{.92}\,TSA_t + \underset{(5.92)}{.90}\,TSW_t - \underset{(4.03)}{38.57}\,NL + \underset{(3.77)}{43.78}\,CONT_t - \underset{(6.17)}{75.64}\,OUT_t, \quad R^2 = .88, \; DF = 38$$

Interpretation of these coefficients is straightforward. Each one point increase in the team slugging average raises the team win percentage by .92 points. A one-hundredth point increase in *TSW* increases *PCTWIN* by .90 points. For equivalent player performance, National League teams will finish nearly 39 points lower in *PCTWIN*, reflecting a higher absolute quality of play. Contenders and cellar teams finish 44 and 76 points, respectively, above or below other teams with equivalent player performance. All of the coefficients are significant at better than the 1 percent level. With an R^2 of .88, the team *PCTWIN* function can be judged as rather completely specified. To the extent that *CONT* and *OUT* do not capture other dimensions of team quality, such as managerial and coaching inputs, the variance associated with omitted factor inputs appears to be quite small.

Team revenue (*REVENUE*) is estimated as a linear equation which specifies *PCTWIN* and market characteristics as the principal determinants.[6] *REVENUE* is defined as home attendance times average ticket price plus revenue from broadcasting rights.[7] The hypothesis is that fan attendance and hence, revenue, is positively affected by team wins. Fans respond to teams that win. When adjusted for differences in monopoly income among the teams, the partial coefficient of *REVENUE* with respect to *PCTWIN* is a measure of marginal revenue across teams. Franchises are granted as exclusive monopoly rights. The size of the geographical area will determine the magnitude of monopoly income. Accordingly, the 1970 population size of the Standard Metro-

[5] Data on *PCTWIN*, *TSA*, *TSW*, and the other variables were obtained from statistical baseball record books.

[6] The linear specification of the revenue function may be surprising, since one would normally expect diminishing returns to victories. However, estimated logarithmically, the coefficient between *REVENUE* and *PCTWIN* was 1.02, which was not significantly different from unity. The fact is that there is no nonlinearity present in the relationships. For computational simplicity in obtaining the player *MRP*, we chose to work with the linear specification. An intuitive justification for the existence of linear revenue functions in baseball is the age and stability of the industry. During the early period, perhaps, increasing returns were present. Currently, there is enough uncertainty of outcome that high win-loss records do not appear to be associated with declining marginal attendance.

[7] Noll kindly made the average ticket price data available. Broadcast revenues were obtained from *Sporting News*, p. 53. Note that the definition of team gross revenue employed here is incomplete. It does not include concession income, which amounts to about forty cents per fan, nor does it subtract the visiting team's share of the gate.

politan Statistical Area (*SMSA*) is included in the revenue function to adjust for the magnitude of monopoly income. Furthermore, interteam differences in *REVENUE* may exist due to interteam differences in the intensity of fan interest in various baseball cities. There is no reason to believe that the marginal revenues of all of the teams are the same. The demand for winning may be different from one baseball city to the next, and this phenomenon may be quite independent of the size of the *SMSA*. To adjust for interteam differences in the intensity of fan interest, the variable *MARGA* is included in the *REVENUE* equation. In the absence of time-series data on average ticket prices and broadcast revenues by team, *MARGA* is the coefficient obtained by estimating team specific attendance equations with time-series data covering the period 1957–71. These team attendance equations essentially estimated the relationship between team percent win and attendance over an extended period holding some other characteristics constant.[8] Wide differences in the slopes were observed, reflecting the phenomenon that teams obtained differential returns in raising their win-loss records. This finding supports Michael Canes' view that there are interteam differences in the elasticity of demand for winning. The partial coefficient of *REVENUE* with respect to *MARGA*, therefore, adjusts marginal revenue across teams for any interteam differences. Three other variables are included in the *REVENUE* function: the *NL* dummy which adjusts for returns due to a higher absolute quality of play in the National League; *STD* which is a dummy variable equal to one for the older stadiums located in poor neighborhoods with limited parking facilities; and *BBPCT*, the percentage of black players on the team. This last variable is included because previous studies have shown the existence of fan racial discrimination in baseball.[9]

The team *REVENUE* function was estimated with team data for 1968 and 1969. The statistical results are:

$$(10)\quad REVENUE_t = \underset{(1.69)}{-\,1{,}735{,}890} + \underset{(6.64)}{10{,}330}\, PCTWIN_t + \underset{(4.61)}{494{,}585}\, SMSA_{70} + \underset{(4.28)}{512}\, MARGA + \underset{(1.84)}{580{,}913}\, NL - \underset{(2.42)}{762{,}248}\, STD_t - \underset{(3.13)}{58{,}523}\, BBPCT_t,$$

$$R^2 = .75,\ DF = 36$$

The empirical results indicate that raising the team win-loss record one point increases team revenue by \$10,330. Since variables have been specified to control for monopoly effects in the industry, this coefficient can be interpreted as a marginal revenue product coefficient free of monopoly effects. For each 1 million *SMSA* population, revenues rise nearly one-half million dollars. For each one point increase in the fan interest coefficient *MARGA*, revenues rise \$512. All of these coefficients are significant at better than the 1 percent level. Membership in the National League is worth nearly \$581,000. The coefficient is significant at the 5 percent level. This income may be due to the higher absolute quality of play

8 In the interest of conserving space, these equations are not presented here. The specification essentially consisted of a relationship between $Attendance_t$ and $PCTWIN_t$ and $PCTWIN_{t-1}$, plus several dummy variables to capture the effects of new stadiums, franchise shifts, and entry of new teams into the same geographical area. $PCTWIN_t$ was almost always highly significant with a coefficient range from 603 to 5,819; $PCTWIN_{t-1}$ was rarely significant. A time-series estimate of team attendance on *PCTWIN* for Seattle was not possible. Therefore, 43 observations were employed in equation (10).

9 See the author (1973, 1974).

in the *NL*. If the team plays in one of the older stadiums, it can expect a reduction in revenue of about $762,000. If this coefficient which is significant at the 1 percent level is reasonably accurate, then it would not appear profitable for the affected team to build privately financed ballparks, even if they serve as multiple sports arenas. Most of the stadiums built during the 1960's cost about $500 per seat or $25 million for a seating capacity of 50,000. The new Three Rivers Stadium in Pittsburgh cost about $700 per seat or $35 million. Thus, servicing the annual debt amounts to about $1.5 to $2 million annually. This factor may explain the domination of publicly owned or subsidized stadia. Finally, a 1 percent increase in the percentage of black players on the team reduces revenues by nearly $59,000. The coefficient is significant at better than the 1 percent level. Overall, 75 percent of the variance in team revenues is associated with the variance in the independent variables specified.

The coefficient of primary interest is *PCTWIN*. From equation (10) a one point increase in *PCTWIN* is estimated to raise *REVENUE* $10,330. From equation (9) a one point increase in *TSA* or *TSW* raises *PCTWIN* by .92 and .90, respectively. Therefore, the marginal revenue product of hitters and pitchers is

$$MRP \text{ hitters} = .92 \times \$10{,}330$$
$$= \$9{,}504 \text{ per point } TSA$$
$$MRP \text{ pitchers} = .90 \times \$10{,}330$$
$$= \$9{,}297 \text{ per } 1/100 \text{ point } TSW$$

However, there are several complications. Since the production function omits a number of inputs, the marginal revenue products of the players may be overstated. Several omitted factor inputs come readily to mind, such as managerial quality, entreprenurial player drafting and trading abilities, and stadium investment. In some crude tests to ascertain the impact of these omitted factors, I found little evidence to suggest any association with *PCTWIN*.[10] Furthermore, the high R^2 in the *PCTWIN* function indicates that the omitted factors do not play a very large role. Nonetheless, certain costs incurred in fielding teams ought to be subtracted from player *MRP* to obtain a measure of player net marginal revenue products. To be conservative, therefore, I will treat these estimates as *gross* marginal revenue products.

To obtain the individual player *MRP*, assumptions have to be made about how individual performance affects team averages. I will assume that individual performance carries with it no externalities, so that team performance is simply the linear summation of individual performance.[11] Most team rosters are divided into ten pitchers and fifteen nonpitchers. Eight of the pitchers will be regular starting or relief pitchers. Therefore, we can

[10] For example, the zero-order correlations between *PCTWIN* and *STD*, $SMSA_{70}$ and *MARGA* were .13, −.03, and −.09, respectively. Therefore, capital investment in stadia or franchise value (correlated with $SMSA_{70}$) has no discernible effect on *PCTWIN*. During the course of the investigation, an attempt was made to specify managerial quality into the *PCTWIN* function. This was approached by estimating the managers' *PCTWIN* record regressed on their teams' *TSA* and *TSW*. It was argued that differences in the slopes of these coefficients would serve as a managerial quality index. These coefficients, however, were not related to current team performance, and, hence, the results are not reported here.

[11] Both player and team externalities may be present in professional baseball. In view of the bitterly competitive environment in professional sports, it is likely that the magnitude of these externalities is small. One player's performance can affect another's, positively or negatively. But there seems no way of incorporating externalities explicitly into the revenue function. Therefore, possibly the estimates of *MRP* for above (below) average players could be biased downward (upward) to some small, unknown degree. For the problem of team externalities, see Canes. Furthermore, in the equation in fn. 14, where individual batting averages are regressed on a number of variables including the rank or order of finish of the team, the coefficient is insignificant. A reasonable interpretation of these results is that team performance does not affect the individual player's performance.

assume reasonably that an average pitcher with a *SW* ratio of about 2.00 will contribute 0.25 (=.125 x 2.00) points to *TSW*. Of the nonpitchers or hitters, about twelve will be regular players or substitutes. Thus, we assume that an average hitter with a *SA* of 340 will contribute 28.3 (=.08333 x 340) points to *TSA*.

Gross marginal revenue products were calculated for players of various talents and are presented in the first column of Table 1. The outstanding characteristics of these estimates are their sheer magnitude. Even mediocre players contribute in excess of $200,000 to team revenues, while star players have *MRP* several times that amount.

How realistic are these *MRP* estimates? Such a question is difficult to answer precisely, since these are the first set of systematically estimated *MRP* for any occupation. However, some comparisons can be made which suggest that these estimates are of the right order of magnitude. Noll recently estimated a statistical demand curve for baseball. In his analysis there are two variables which capture the player's contribution to team attendance. One variable is the team's number of games out of first place. This variable is akin to the *PCTWIN* variable in equation (9). The second variable is a superstar variable, which Noll interprets as capturing the effect of superstars on attendance independent of their contribution to team victories. Neither variable is statistically convincing.[12] Statistical significance aside, some useful comparisons can be made. Taking Hank Aaron, who in 1968 had a slugging average of 498 and 11 percent of the team at bats, arbitrarily as an example, I calculate Aaron to be worth $520,800. Using Noll's coefficients, in an average size baseball city, Aaron was worth $225,000 because he was one of two players with superstar status, plus $375,000, because he would be the difference between victory and defeat in about 20 games. Thus, by Noll's method, in 1971 Aaron was worth about $600,000. Noll's results would seem to confirm the order of magnitude of the *MRP* in Table 1.

As a second comparison, David Davenport by a crude procedure estimated Sandy Koufax's *MRP* at $617,554. Using equation (9), dividing the *PCTWIN* coefficient in equation (10) by the 1969 average ticket price to obtain marginal attendance, multiplying by the 1966 average ticket price, and noting that Koufax had a *SW* ratio of 4.12 in 1966, his last season, and pitched 22.2 percent of the team's innings, I estimate that Koufax was worth about $725,000.

To obtain player net marginal revenue products, the compensation to other inputs and firm specific training or player development costs need to be subtracted. Unfortunately, this cannot be done precisely because of data limitations. However, it is possible to crudely measure average costs in baseball and subtract them from player marginal revenue products. In view of the fact that little correlation was discovered between *PCTWIN* and a number of nonplayer inputs, this may be an appropriate procedure a priori. In undertaking this procedure, I have been guided by two principles. First, I chose to overstate other factor costs so that the net marginal revenue products may be viewed as a lower bound. Secondly, where stated costs, such as administrative expenses, appeared to reflect attempts to hide profits, I chose to work with the smaller cost figures. These factors will become clear in the discussion.

In the course of the Curt Flood litiga-

[12] The t-values for Noll's games behind variable was 1.36, while the t-value for the superstar variable was 1.8 (see equation (7)). Since there is a great affinity between the games behind variable and the *PCTWIN* variable, Noll's relatively poor results can be reasonably interpreted as caused by collinearity between team wins and the number of superstars.

tion, the industry presented testimony on some aspects of its costs. Noll utilized these data in his study on the profitability of sports franchises. Using Noll's data, four general categories of costs can be identified: 1) team costs, which include principally roster and team specific nonplayer salaries; 2) game costs, which include transportation costs for eighty-one away games, equipment, stadium rental, etc.; 3) general administrative costs, which are principally salaries of the front office personnel; and 4) sales costs. The average team in 1969 had team costs of $1,072 million. The average player salary in 1969 was $25,000 for a total average roster cost of $625,000. Distributed over the twenty-five-man roster, this yields nonplayer team costs of about $14,000 per player. Minimum game costs were $800,000. These costs ought to be spread over the principal team personnel, that is, players, coaches, and managers. Dividing conservatively by 30, per player game costs amount to $26,700. Similar procedures for administrative and sales costs yielded per player costs of $13,300 and $3,300, respectively. Total nonplayer costs per player amount to $57,300. This leaves two costs to be accounted for: capital and training. Capital investment in sports franchises is an elusive concept. Classically, we think of capital investment as plant and equipment necessary for the production process. Capital investment in sports franchises consists of the purchase of a monopoly right and player contracts—a monopsony right. In some cases, stadiums are owned but more frequently they are leased. Controversy aside as to what constitutes capital in the industry and whether this has anything to do with the productive process, the average value of a baseball franchise is $8.4 million. Figuring an opportunity cost of 10 percent, capital costs spread over the team amount to $28,000 per player. Finally, according to industry reports, the average cost of player development is about $300,000 per player. This estimate is derived by dividing the entire cost of minor league operations by the number of players promoted to the major leagues. Such an estimate of training costs is a gross overstatement. However, taking the figure at face value, this amounts to an average annual cost of about $43,000 per player (average player life is seven years). Therefore, figured in this way, the total annual cost of fielding an average career length player amounts to about $128,300 plus his salary.

TABLE 1—ESTIMATED MARGINAL REVENUE PRODUCTS AND SALARIES OF AVERAGE CAREER LENGTH BASEBALL PLAYERS

Performance $\overline{SA}$ or $\overline{SW}$	Gross Marginal Revenue Product	Net Marginal Revenue Product	Salary
	Hitters		
270	213,800	85,500	31,700
290	230,000	101,700	34,200
310	245,200	116,900	36,800
330	261,400	133,100	39,300
350	277,500	149,200	41,900
370	292,700	164,400	44,400
390	308,900	180,600	47,000
410	325,000	196,700	49,600
430	340,200	211,900	52,200
450	356,400	228,100	54,800
470	372,600	244,300	57,400
490	387,800	259,500	60,000
510	403,900	275,600	62,700
530	420,100	291,800	65,300
550	435,300	307,000	67,900
570	451,400	323,100	70,600
	Pitchers		
1.60	185,900	57,600	31,100
1.80	209,200	80,900	34,200
2.00	232,400	104,100	37,200
2.20	255,700	127,400	40,200
2.40	278,900	150,600	43,100
2.60	302,200	173,900	46,000
2.80	325,400	197,100	48,800
3.00	348,600	220,300	51,600
3.20	371,900	243,600	54,400
3.40	395,100	266,800	57,100
3.60	418,400	290,100	59,800

The net marginal revenue products for players with an average career length are

presented in the second column of Table 1. Subtracting such costs substantially reduces the gross marginal revenue products of the players. Note, however, that even mediocre players with an average career length have net marginal revenue products in excess of their estimated salaries. Clearly, if the per player training investment cost of $300,000 is accepted at face value, mediocre players at some point less than the average career length will have negative net marginal revenue products. Two final points are worth making. First, training investment may vary with the quality of the player. Second, pitchers may require a greater training investment than hitters. My suspicion is that mediocre players require more training than superior players. Thus, training costs may be understated for the below average and overstated for the above average players. I also suspect that training costs are overstated for the hitters and understated for the pitchers.

IV. Salary Determination

In order to test the salary model outlined, data are required on player salaries and performance. Baseball, unlike other sports where player salary information is guarded, publicizes the salaries of its ball players. Frequently, the press is used as a bargaining tool in salary negotiations by communicating inside information on player demands and owner contract offers. Thus, annually there is widespread public information on individual player salaries. Additionally, there is a wealth of performance data on teams and players. Baseball statistics on player and team performance are the best of any sport.

Furthermore, baseball owners and players alike agree that salary and player performance are related and that playing performance is measurable. However, there is disagreement over which measures accurately reflect performance. Part of this controversy is due to the fact that baseball simultaneously is an offensive and defensive team sport. Disagreements about performance may partly depend on which aspect of the game is thought more important. The variables appearing in the regressions below are thought to be the best single, independent measures of offensive and defensive player performance available. The variables were not selected on entirely a priori grounds. Rather, experimentation with every known measure of performance was undertaken and then these performance measures were selected.[13]

Statistically, four factors seem important in the determination of major league

[13] I was guided initially in the specification of the player performance measures for the salary equations by the previous work of Pascal and Rapping. For hitters, they tried lifetime batting average, home runs, times at bat last season, age, and education. The performance measures for pitchers were games won per full season played, innings pitched per full major league season, total major league games won, total major league innings pitched, total seasons pitched, and differences between performance last year and lifetime performance. In one form or another, these variables explained a considerable portion of the variance in players' salaries (see their Table 3). Additionally, for nonpitchers, I calculated such measures as total bases, runs produced, extra base power, best at getting on base, base stealing, and fielding average. Such measures as pitching percentage, earned run average, hits divided by total batters faced, and games completed divided by games started were tried for pitchers. Furthermore, I considered certain factors not explicitly related to past performance. Since salary negotiations are based on expected performance, there is a certain risk associated with the prediction. Past deviations from average performance seemed like an appropriate way of incorporating expectations into the salary model. I also believed that required performance levels were likely to differ position by position, at least in the means, and attempted to take such differences into account along with the interclub average salary differentials. I believe no important factor was left uninvestigated. The fact that one performance measure or another or one plausible effect or another does not appear in the regression equations reported here does not mean that the measure or the effect was not associated with salary variations. Quite the contrary, most of the effects described above frequently were highly correlated with players' salaries, but their unique effects could not be isolated. For example, batting averages lagged one period and percentage deviation of the lifetime batting averages were significantly related to hitters' salaries but not independent of lifetime batting average.

ball player salaries. These factors are: hitting or pitching performance; the weight of the players' contributions to team performance; the number of years spent in the majors; and the greater bargaining power of "star" or "superstar" players.

One factor which governed the selection of the performance measures was that they be independent measures of the player's contribution to team performance. For hitters, the performance measure selected was lifetime slugging average, $\overline{SA}$. The lifetime batting average, which is a more popular measure of performance, does not take into account the number of bases advanced on a hit, although number of bases advanced is critical in scoring runs and hence winning games. Admittedly, some excellent hitters have low slugging averages. This difficulty was overcome by specifying a dummy variable $D_{\overline{BA}}$ in the hitter's salary equation which was equal to unity for those players with below average lifetime slugging averages but above average lifetime batting averages. The weight of the hitter's contribution $\overline{AB}$ is specified as his total lifetime at bats divided by his number of years in the majors times 5,500, the average season at bats for a major league team. The variable $\overline{AB}$ is a measure of the player's percentage contribution to team performance.

Pitching performance is more difficult to measure. The most common measure of pitching performance is the earned run average. However, the measure is virtually uncorrelated with salary. Pascal and Rapping found that the win-loss record was related to pitcher salary. However, neither measure of pitching performance is independent of team performance. Neither measure uniquely measures the pitcher's contribution to the team. In this study, the lifetime strikeout-to-walk ratio $\overline{SW}$ is employed as the single best measure of pitching performance. I also specify the lifetime average percentage of innings pitched out of total innings $\overline{IP}$ (approximately nine x games x years), as another important dimension of the pitcher's contribution.

Player salaries appear to rise automatically over time. Therefore, years spent in the majors M affect salary independently of average lifetime performance. This independent effect of time may be a pure seniority effect or it may measure the separate contribution of experience to playing ability. However, M seems to measure the phase in the career of the ballplayer and hence may adjust performance for trend.[14]

In view of the fact that size of $SMSA$, intensity of fan interest, and membership in the National League affect team revenues, one might suspect an influence of these variables on player salaries. It is a common notion that players are more highly paid in the National League and in the larger $SMSA$s. Economic theory cannot answer this question a priori. On the one hand, there is the view that the strong monopsony power of a monopolist located in a large market would yield lower salaries for such teams. Consistent with this view would be the greater demand of players for location in the large $SMSA$. Players derive nonbaseball income from

[14] This point is discussed in more detail in the author (1973). To determine whether the M coefficient largely reflected productivity increases over time, career batting averages for individual major league outfielders on teams in 1971 were estimated with the following statistical results:

$$BA_{it} = \underset{(.94)}{12.99} + \underset{(.63)}{1.483}\,NL - \underset{(.47)}{.1863}\,RANK + \underset{(17.45)}{.8850}\,\overline{BA} + \underset{(4.50)}{9.030}\,Log\,M,$$

$$R^2 = .42,\ DF = 725$$

The specification argues that the ith individual's batting average in season t is determined essentially by his lifetime career batting average or the mean batting average and the length of time in the majors. Neither NL nor the order of the finish of the team have an independent effect on BA_{it}. The results suggest about a 9 point increment in BA per season. Converted to an elasticity at the mean, productivity increases about 3.5 percent per season.

advertising firms and other sources which are located primarily in the large coastal population centers. On the other hand there is the view that within a bargaining framework, labor (players) would share in monopoly profits. Hence, this view would predict higher salaries where monopoly profits are large. As an interesting empirical test, $SMSA_{70}$, $MARGA$, and NL are included in the salary regressions.

Finally, players of rare talent, the stars and the superstars, command salaries apparently in excess of their relative contribution to the team. This introduces a nonlinearity into the salary equations. The nonlinearity may be explained by the greater bargaining strength of these players. There are very few substitutes for star players. Alternatively, the marginal revenue products of star players may be higher than their performance would strictly warrant. While I have not been able to confirm the fact empirically, fans may be drawn to the ballpark on the margin to see these players independent of their performance. To account for this effect, the salary regressions are estimated logarithmically.

There were two primary sources for the data used in the salary equations. All of the values of the independent variables were obtained or calculated from or based on *The Baseball Encyclopedia*. The salary data were made available to me by Pascal and Rapping and hence is the same salary sample as used in their study. Their data were obtained from reported salaries in twenty local newspapers and *The Sporting News*. There is reason to believe that the data are fairly accurate, but, equally important, there is no a priori reason to think that reporting errors, if any, are systematically related to any of the variables used in the analysis.[15]

The salary sample is nonrandom and biased toward the upper tail of the salary distribution. There are 148 observations, 87 of which are for 1968 salaries and the remainder for 1969. Of the 61 observations for 1969, 41 of the players also appear in the 1968 sample. Therefore, there are 107 nonrepeat observations. The salary range in the observations is from \$10,000 to \$125,000 with 34 observations below \$25,000. Thus, about 25 percent of the sample is in the lower end of the salary observations. In our judgment, enough observations over the whole salary range are present to make inferences about the populations. The mean salary of all the players in the sample was \$48,100. This figure may appear high, but it does not compare unfavorably to the reported average salary of \$42,200 in 1970 of 209 veterans on the major league teams.[16]

The player salary regressions are as follows:[17]

$$
\begin{aligned}
(11)\quad Log\, S_{hitters} = \underset{(.82)}{.6699} &+ \underset{(4.76)}{1.0716}\, Log\, \overline{SA} \\
&+ \underset{(7.53)}{.5220}\, Log\, M + \underset{(1.61)}{.0579}\, Log\, D_{\overline{BA}} \\
&+ \underset{(3.10)}{.2746}\, Log\, \overline{AB} \\
&- \underset{(.78)}{.0621}\, Log\, SMSA_{70} \\
&+ \underset{(2.81)}{.2645}\, Log\, MARGA \\
&+ \underset{(.62)}{.0194}\, Log\, NL, \\
&\qquad R^2 = .81,\ DF = 85
\end{aligned}
$$

[15] Pascal and Rapping verified the salary data with Marvin Miller, head of the Major League Baseball Player's Association. For further discussion of the various data, see Pascal and Rapping and the author (1974).

[16] See the testimony of John Clark, Jr., Curt Flood Court Transcript, p. 1660.

[17] Since current salary depends on last year's lifetime performance, all of the independent variables were lagged one year. The t-values appear under the regression coefficients and *log* values are to the base 10.

$$(12)\quad Log\, S_{pitchers} = \underset{(4.16)}{3.3845} + \underset{(4.00)}{.8076}\, Log\, \overline{SW} + \underset{(7.70)}{.5015}\, Log\, M + \underset{(4.22)}{.9698}\, Log\, \overline{IP} - \underset{(.53)}{.0534}\, Log\, SMSA_{70} - \underset{(.61)}{.0619}\, Log\, MARGA - \underset{(.20)}{.0070}\, Log\, NL,$$

$$R^2 = .78, \; DF = 48$$

The main determinants of player salaries are, except for one case, $D_{\overline{BA}}$, which is significant at the 10 percent level, all significant at the 1 percent level or better. Overall, 78 to 81 percent of the variation in ballplayer salaries is accounted for. The remaining variance may reasonably be attributed to the vagaries of the bargaining process. However, as can be readily seen, the salary process in major league baseball is quite deterministic. The concept that ballplayer salaries are related to performance seems reasonably well confirmed.

With regard to the effect of $SMSA_{70}$, $MARGA$, and NL, it can readily be ascertained from the equations that while these variables play a substantial role in revenue determination for teams, they have virtually no effect on player salaries. Particularly noteworthy is that the $SMSA_{70}$ coefficient is in fact negative but in both cases insignificant. The $SMSA$ size has no effect on player salaries. While the absolute quality of play is higher in the National League, this effect does not raise player salaries. Finally, intensity of fan interest, $MARGA$, does not affect pitchers' salaries, but does contribute positively and significantly to hitters' salaries.

In the fourth column of Table 1 the predicted salaries of the ballplayers are reported. In determining the predicted salary, the variables other than the performance measures were set equal to their respective population means. For hitters, M was equal to 8 in the sample, while pitchers spent about six years in the majors. The shorter playing careers of pitchers is well known. On the average, pitchers have about 14 percent of the total innings pitched during a regular season. This is equivalent to about 207 innings or 23 game equivalents. Average hitters have about 7 percent of the at bats. Above average hitters are paid somewhat more for their performance, *ceteris paribus*, than above average pitchers. Undoubtedly, this reflects the effect of a two year average difference in time spent in the majors. However, below average hitters appear to earn less than below average pitchers. In part, this may be due to the fact that as good starting pitchers age, they can be used effectively by starting them less frequently and using their skills in relief. On the other hand, effective hitting cannot be improved by resting the batter.

V. The Degree of Monopsonistic Exploitation of Professional Ball Players

Once knowledge of player marginal revenue product and salary is obtained, it is a relatively simple matter to obtain estimates of the degree of monopsonistic exploitation. In Table 2, gross and net marginal revenue products, salary, and rates of monopsonistic exploitation are presented for three categories of players. Mediocre players are assumed to have career lengths of four years, with hitters having 5.0 percent of the team at bats, AB, and pitchers having 10.0 percent of the innings pitched, IP. Average players are assumed to have career lengths of seven years, with $AB=8.0$ percent and $IP=14.0$ percent assumed. Star players are assumed to have career lengths of ten years, with $AB=10.0$ percent and $IP=18.0$ percent. On the basis of an analysis of hitters'

Table 2—Marginal Revenue Product, Salary and Rates of Monopsonistic Exploitation for Three Quality Categories of Players

Career Performance $\overline{SA}$ or $\overline{SW}$	Gross Marginal Revenue Product	Net Marginal Revenue Product	Salary	Rate of Monopsonistic Exploitation $(GMRP-S)/GMRP$ Columns (2–4)/2	Rate of Monopsonistic Exploitation $(NMRP-S)/NMRP$ Columns (3–4)/3
Mediocre Hitter ($M=4$, $AB=5.0$ percent)					
255	121,200	−39,100	9,700		
264	125,500	−34,800	13,900		
273	130,200	−30,100	17,200		
283	135,000	−25,300	20,000		
Total	511,900	−129,300	60,800	.88	1.47
Mediocre Pitcher ($M=4$, $IP=10.0$ percent)					
1.50	139,500	−20,800	9,000		
1.55	144,100	−16,200	12,800		
1.61	149,700	−10,600	15,700		
1.66	154,300	−6,000	18,100		
Total	587,600	−53,600	54,800	.91	2.02
Average Hitter ($M=7$, $AB=8.0$ percent)					
305	231,900	103,600	14,100		
316	240,500	112,200	20,300		
327	249,000	120,700	25,000		
338	256,600	128,300	29,100		
350	266,100	137,800	32,700		
362	275,600	147,300	36,000		
375	285,100	156,800	39,000		
Total	1,804,800	906,700	196,200	.89	.79
Average Pitcher ($M=7$, $IP=14.0$ percent)					
2.00	260,300	132,000	16,500		
2.07	269,600	141,300	23,300		
2.14	278,900	150,600	28,600		
2.22	288,200	159,900	33,000		
2.30	297,500	169,200	36,900		
2.38	306,800	178,500	40,500		
2.46	316,000	187,700	43,700		
Total	2,017,300	1,119,200	222,500	.89	.80
Star Hitter ($M=10$, $AB=10.0$ percent)					
385	365,900	250,600	20,400		
398	378,300	263,000	29,400		
412	391,600	276,300	36,300		
427	405,800	290,500	42,200		
442	420,100	304,800	47,400		
457	434,300	319,000	52,100		
473	449,500	334,200	56,500		
490	465,700	350,400	60,500		
507	481,900	366,600	64,400		
525	499,000	383,700	68,000		
Total	4,292,100	3,139,100	477,200	.89	.85
Star Pitcher ($M=10$, $IP=18.0$ percent)					
2.60	437,000	321,700	27,200		
2.69	446,300	331,000	38,500		
2.79	464,900	349,600	47,200		
2.88	483,400	368,100	54,500		
2.98	502,000	387,500	61,000		
3.09	520,600	405,300	66,800		
3.20	539,200	423,900	72,100		
3.31	557,800	442,500	77,100		
3.42	576,400	461,100	81,800		
3.54	595,000	479,700	86,300		
Total	5,122,600	3,969,600	612,500	.88	.85

career performance, all players were assumed to increment their performance at the rate of 3.5 percent per season.[18] The principal difference in the marginal revenue products of the categories of players is due to higher lifetime performance levels and greater utilization of the better players. The principal difference in the net marginal revenue products arises from the higher gross marginal revenue products of the better, more intensively used players and the longer career length over which training costs are recovered. The rate of monopsonistic exploitation in the fifth column, which relates salary to the gross marginal revenue product, may be viewed as the upper bound of player exploitation, while the sixth column, which relates salary to the net marginal revenue product, may be viewed as the lower bound.

As examination of Table 2 reveals, players are generally exploited to a large degree in major league baseball. The exploitation index is bounded between 0 and 1 when $MRP>0$. A value of 1.0 indicates complete exploitation and zero indicates no exploitation. Considered over career length, average players receive salaries equal to about 11 percent of their gross and about 20 percent of their net marginal revenue products. Star players receive about 15 percent of their net marginal revenue products. Only mediocre players have salaries in excess of their net marginal revenue products. On the whole, therefore, it seems that the economic loss to professional ballplayers under the reserve clause is of a considerable magnitude.

VI. Policy Implications

Economic analysis points to the exploitation of the professional baseball player under the reserve clause through the introduction of monopsony power. Empirical analysis confirms the existence of this exploitation and suggests that it is of considerable magnitude. The exemption of organized baseball from the antitrust laws has created such player economic rents. The decision of the Supreme Court against Curt Flood by no means removes the issue. The players are increasingly militant over the reserve clause and have made known their desire for modification of the clause in their collective bargaining negotiations. In the last round of negotiation, players obtained a concession of binding arbitration of salary disputes, but player-initiated transfers still are not possible. Furthermore, Congress is currently studying the problem and some alternatives may emerge. I can only mention and comment on the implications of some of the major suggestions.

One proposal would place baseball players under a contractual arrangement like pro football. Under this system, known as the "Rozelle Rule," a player can play out his option and become a free agent, but since the team purchasing the player must come to terms with the team for whom the player previously played, player-initiated transfers are an illusion. A similar idea in the proposed *NBA-ABA* merger agreement would produce similar consequences. Under their proposal a player can submit a salary dispute to nonbinding arbitration. If still dissatisfied, the player can switch to another team only if the two teams agree on a transfer price.

A second popular proposal revolves around a system of long-term contracts. Here players would be free agents until they sign a contract for five years or so. Their free agent status would revert at the expiration of the contract. Such a system would partially redistribute economic rents, would recover team player development costs, and would have minimal effect

[18] See fn. 14. Treating all players, both hitters and pitchers, over the quality range of performance as incrementing their performance at the rate of 3.5 percent per season is arbitrary.

on the equalization of team playing strengths.

The most radical proposal is a completely free labor market with all contracts for a full season being negotiated off-season. This proposal would eliminate player economic rents. Organized baseball argues that such a scheme would destroy the game. They point to the rich owner, who could not be prevented from buying all of the good players. They argue that investments in teams would be unattractive. Teams would fold and no buyers would be found. They also forecast the end of player development and minor league subsidies and hence long-term damage to the sport. None of these objections has more than superficial merit. As El Hodiri and Quirk have proved, the distribution of playing talent would be the same with or without the reserve clause; since the *MRP*s of players vary by city size, only equal distribution of revenues among the teams would effectively bring about equalization of play. Profitable investments in teams are still possible in the absence of monopsony rents, since monopoly profits remain. It is possible that reorganized with the league as the basic economic decision unit, instead of the team, baseball could operate as an efficient monopoly or duopoly rather than as an inefficient cartel. Finally, if players were to receive their *MRP*s, the players as in the Professional Golf Association would fully absorb their training costs. Baseball schools could develop players with the potential major leaguer paying tuition. These schools could be within organized baseball or independent of it. Or, perhaps, colleges will subsidize player development costs in baseball as they do in football and basketball. With modification in the reserve clause, the players could benefit and the game need not suffer.

REFERENCES

M. E. Canes, "Public Policy Toward Professional Team Sport," in R. G. Noll, ed., *Government and the Sports Business*, Washington 1974.

J. Clark, Jr., *Testimony*, Curt Flood Court Transcript, Flood vs. Kuhn, 407 U.S. 258, Washington 1972.

D. S. Davenport, "Collusive Competition in Major League Baseball, Its Theory and Institutional Development," *Amer. Econ.*, Fall 1969, *14*, 6–30.

M. El Hodiri and J. Quirk, "An Economic Model of a Professional Sports League," *J. Polit. Econ.*, Nov./Dec. 1971, *79*, 1302–19.

I. Horowitz, "Professional Sports Broadcasting and the Promotion of Sequential Oligopoly," in R. G. Noll, ed., *Government and the Sports Business*, Washington 1974.

R. G. Noll, "Attendance, Prices, and Profits in Professional Sports," in his *Government and the Sports Business*, Washington 1974.

A. H. Pascal and L. A. Rapping, "The Economics of Racial Discrimination in Organized Baseball," in A. H. Pascal, ed., *Racial Discrimination in Economic Life*, Lexington 1972, 119–56.

G. W. Scully, "The Economics of Discrimination in Professional Sports: The Case of Baseball," in R. G. Noll, ed., *Government and the Sports Business*, Washington 1974.

———, "Economic Discrimination in Professional Sports," *J. Law Contemporary Probl.*, Winter-Spring 1973, *38*, 67–84.

The Baseball Encyclopedia, New York 1969.

Sporting News, Mar. 18, 1972.

Competitive Interest Payments on Bank Deposits and the Long-Run Demand for Money

By BENJAMIN KLEIN*

This paper tests two assumptions made in previous demand for money studies. The first assumption concerns the role assigned to "the" interest rate in the demand for money function. Common practice identifies the rate of interest with "the opportunity cost of holding money." In Section I, this formulation is shown to blur an important distinction between "the price of money" and "the price of money substitutes" and to implicitly assume that the relevant price variable is the difference between these two distinct prices.

A second assumption is that the current ban on interest payments on demand deposits is fully effective. In Section II, I operationally define these price variables and estimate perfectly competitive interest payments on commercial bank deposits. This is done by crudely measuring commercial bank marginal costs and assuming that all "excess" profit is passed on to depositors in indirect ways. Section III presents estimates of a more complete long-run demand for money relationship which explicitly includes an own price and a cross price as separate influences, and assumes that the prohibition of the payment of interest on deposits is totally ineffective. Although no firm conclusion is reached concerning the proper functional specification of the demand function, the results significantly improve upon previous demand for money estimates which assumed that the interest payment prohibition is totally effective. In Section IV, the results are summarized and implications of the analysis for monetary theory are presented. My competitive demand for money formulation leads to a reinterpretation of the existing evidence on what is generally referred to as "the interest elasticity of demand for money." This estimate is shown to be significantly biased downward. It also implies that the demand for and supply of money are more interdependent than is usually assumed.

I. The Own and Cross Prices of Money

Although theoretical work on the microfoundations of monetary exchange is progressing rapidly, money is still commonly considered empirically to be a durable producer and consumer good which yields an unspecified "monetary service" flow. This study, like most previous empirical demand for money studies, ignores the important unanswered theoretical question of what these services consist of. The services are merely assumed to enter a utility function and the demand for money is then implicitly derived from the demand

* University of California, Los Angeles. This paper is based on my unpublished Ph.D. dissertation. I am indebted to Milton Friedman, Levis Kochin, Anna Schwartz, and the managing editor of this *Review* for useful comments and to Irene Abramson, Stephen Ferris, and Charles Lieberman for able research assistance. I am also grateful to Friedman and Schwartz for graciously supplying much of the underlying data employed in this study and to the UCLA Institute of Government and Public Affairs and the National Bureau of Economic Research for providing financial support. This paper is not, however, an official National Bureau publication since it has not undergone the full critical review accorded National Bureau studies.

for these unspecified monetary services.[1] The purpose of the following exercise is not to throw light upon the analytical foundation of the demand for money, but by deriving a demand function, to properly define price of money variables.

Assume that in addition to money there exists one other dollar denominated financial asset which is an alternative source of monetary services. The flow of monetary services yielded an individual can then be represented by:

$$(1) \qquad N = N(M/P, S/P, \beta)$$

where N is the flow of real monetary services yielded per unit time, M/P is the stock of real cash balances held, S/P is the real stock of the monetary substitute asset held, and β is a portmanteau variable representing all other possible N-determining variables which are assumed to remain constant throughout this analysis. Equation (1) can be thought of as the individual's production function for monetary services with both (M/P) and (S/P) assumed to have declining marginal productivities.

For simplicity and to facilitate comparisons between financial assets and commodities, let X represent a scalar measure of a vector of the commodity service flows yielded by the nonfinancial goods and services consumed by the individual. The price level P is then defined as the dollar rental price of a unit rate of commodity services, measured in dollars per unit time per X. The individual's utility function can then be written:

$$(2) \qquad U = U(X, N)$$

where X equals the rate of consumption of commodity services and N equals the rate of consumption of monetary services.

Assume that in addition to yielding marginal nonpecuniary monetary service returns, money yields a marginal pecuniary interest rate return equal to r_M and the substitute asset yields a marginal pecuniary interest rate return equal to r_s. Assume further that there exists another financial asset called a bond B, which yields no nonpecuniary monetary service returns and marginal pecuniary interest payments equal to i.

Finally, consider the individual's rate of consumption of commodity services. The X flow that can be purchased and consumed per unit time by the individual is a function of the rate of net dollar receipts to the individual.

$$(3) \quad X = [Py_0 + r_M \cdot M + r_s \cdot S + iB](1/P)$$

where X equals the rate of flow of commodity services; y_0 equals the individual's given real rate of permanent earnings, measured in commodity services; M, S, and B equal the dollar value of the individual's money, substitute asset and bond holdings, measured in dollars; and r_M, r_s and i are the respective rates of return on these financial assets, measured in dollars per dollar per unit time. The term in brackets represents the individual's money income.

Now assume the individual has a given total of real nonhuman wealth W_0 where

$$(4) \qquad W_0 = M/P + S/P + B/P$$

and the individual maximizes utility subject to his given real human and nonhuman wealth (y_0 and W_0) and to the budget constraint that all income is spent. From (2) and (3) form the Lagrangian

$$(5) \quad V = U(X, N) - \lambda[PX - Py_0 - r_M \cdot M - r_s \cdot S - iB]$$

From (4) note that

$$(4') \qquad B = PW_0 - M - S$$

If we assume that the individual takes the market interest yields and the rental price

[1] See for example, Milton Friedman (1956), p. 4 and Gregory Chow, p. 113. For a more complete analysis and an attempt to partially standardize the money stock in terms of its service flow, see the author (1972).

of output services as a constant with respect to his own purchases, then substituting (4′) into (5) and differentiating with respect to the individual's decision variables X, M, and S, the necessary conditions for a constrained maximum are:

$$(6)\quad \frac{\partial V}{\partial X} = U_X - \lambda P = 0$$

$$(7)\quad \frac{\partial V}{\partial M} = U_N N_{(M/P)}(1/P) - \lambda(-r_M + i) = 0$$

$$(8)\quad \frac{\partial V}{\partial S} = U_N N_{(S/P)}(1/P) - \lambda(-r_s + i) = 0$$

Assuming the usual second-order conditions hold, the optimum quantities of X, M, and S are then given where

$$(9)\quad \lambda = \frac{U_X}{P} = \frac{U_N N_{(M/P)}(1/P)}{(i - r_M)} = \frac{U_N N_{(S/P)}(1/P)}{(i - r_s)}$$

Equation (9) states the usual equilibrium condition that the marginal utility of a dollar of income spent in all directions must be equal; i.e., marginal utility divided by price must be the same for all goods.[2] The rental price of a unit of commodity services is P while $(i - r_M)$ denoted by P_M is the rental price of the monetary service stream from a dollar of money and $(i - r_s)$ denoted by P_S is the rental price paid for the monetary service stream from a dollar of monetary substitutes.[3] The rental price $(i - r_M)$ can be considered as the marginal pecuniary alternative cost per unit time of holding a dollar of money and in equilibrium will equal the value of the monetary services from a marginal dollar of money.[4] Similarly, $(i - r_s)$ represents the marginal alternative cost of holding the monetary substitute asset and in equilibrium will equal the value of the monetary services from a marginal dollar of substitute asset.

If we assume that permanent income y_p can be considered as the relevant empirical constraint (i.e., proxy for y_0 and W_0), then the conditions of utility maximixation imply that the demand for real cash balances may be written as:

$$(10)\qquad (M/P)^d = f(y_p, P_M, P_s)$$

where $\partial(M/P)^d/\partial y_P > 0$, $\partial(M/P)^d/\partial P_M < 0$ and, as long as the substitution in production effect dominates any scale of production effect,[5] $\partial(M/P)^d/\partial P_s > 0$. The variable P_M can be considered as the own price and P_s as the cross price of money and both P_M and P_s should be entered as variables in the demand for money function.

Equation (10) does not resemble common formulations of the demand for money which are typically represented by:

$$(11)\qquad (M/P)^d = h(y_p, r_s)$$

Some sense can be made of formulations such as (11) if we interpret them as im-

[2] There may be some confusion with regard to the good "money" since λ is sometimes called "the marginal utility of money." This is incorrect; λ should more properly be called the marginal utility of money income (i.e., the change in an individual's utility from a one dollar change in his income allocated optimally across goods) and distinguished from the marginal utility of money U_M (i.e., the change in an individual's utility from a dollar change in his money holdings). Equation (9) states that in equilibrium U_M and λ are related by the income foregone in holding a dollar of money $(i - r_M) = U_M/\lambda$.

[3] Edgar Feige (1964) defines similar "prices" but in his empirical analysis assumes that the pecuniary rate of return on the asset yielding no nonpecuniary monetary services remains fixed, leaving alternative r_s measures and r_m as the sole variables in the demand functions.

[4] Alternatively, interest payments plus the value of monetary services equals i. When net pecuniary and nonpecuniary returns are considered, then money and every other asset traded in the economy must in equilibrium yield on the margin *the* interest rate.

[5] An increase in P_s will, *ceteris paribus*, increase the demand for money for a given desired monetary service flow; but an increase in P_s will decrease the monetary service flow demanded and hence decrease the demand for money (see the author (1972)). The first effect is assumed to always dominate and our alternative asset is assumed to be a substitute for money in the sense of a positive cross partial derivative.

plicitly assuming that (a) the relevant price variable in the demand for money function is the difference between the own and cross prices of money, and (b) interest payments on money are zero. Assumption (a) implies that i, the benchmark return on the asset yielding no monetary services, need not be considered in the demand function; and assumption (b) implies that the rate of interest paid on money need not be considered.

Most previous demand for money estimates have made these two assumptions. Assumption (b) is generally recognized as a deficiency in existing demand for money studies. David Laidler (1966b, p. 545, fn. 4) explains this discrepancy between what is recognized to be theoretically correct and existing empirical work on the grounds that reliable data on interest payments on deposits over a substantial time period do not exist. Explicit interest payments were not systematically reported until 1919 and implicit interest payments have never been systematically reported.[6] The latter fact is of particular importance with regard to demand deposits where explicit interest payments have been prohibited since 1933. The two previous demand for money studies that have not assumed that interest payments on money are zero (Richard Selden and Tong Hun Lee (1967)) cover much shorter time periods (1919–51 for Selden and 1951–65 for Lee) and measure interest on demand deposits post-1933 as the negative of bank service charges, thereby assuming that the interest prohibition has been totally effective and ignoring implicit interest payments completely. This study takes account of implicit interest payments on deposits by making the alternative assumption that the interest prohibition is totally ineffective.[7]

On the other hand, assumption (a) has not generally been recognized as a deficiency in existing demand for money studies. Both Selden and Lee, for example, assume that $(P_M - P_s)$ is the correct variable by considering the relevant price variable to be the difference between the yield on money substitutes and the yield on money, $(r_s - r_M)$. This "net interest rate" variable confuses the own and cross price effects and makes it difficult to interpret their empirical results. The separate response of the demand for money to each price cannot be determined, and the specification implicitly assumes that changes in the benchmark rate of return have no effect on the demand for money.[8] If we

[6] Implicit interest payments may take many forms. Reduced interest rates on loans to depositors combined with compensating balance requirements may be a major avenue by which implicit payments are made. The compensating balance requirement is then considered not an added constraint to the consumer but merely ties the loan rate explicitly to "working" deposit holdings. This may explain why the arrangement generally exists on business rather than consumer loans (those customers most likely to receive interest payments on their deposits) and why the requirement is always stated in terms of average rather than minimum deposit levels. It may also explain why stated commercial bank lending (say, prime) rates, which are really rates net of implicit interest payments, can be below market borrowing (say CD) rates. (See the author (1970, Appendix 1) for a discussion of the deposit-loan reciprocity arrangement between commercial banks and their customers.) In this paper, however, I am not concerned with how commercial banks may evade maximum deposit interest regulations but solely with whether the assumption of complete evasion explains observable phenomena better than the alternative assumption that these regulations are never evaded.

[7] Michael Darby has recently used a crude version of the r_M series I calculate here in quarterly demand for money regressions for the postwar period with results favorable to my hypothesis. (He, however, uses total reserves rather than the correct concept of reserves held by banks against demand deposits and also fails to make any adjustment for possible returns from the Federal Reserve to commercial banks in deriving his r_{M_1} series.) In addition, Robert Barro and Anthony Santomero have recently used the results of a private survey of commercial banks concerning the reported rate at which service charges are remitted as a function of demand deposit balances on small consumer-type accounts in a household demand for money function. The correlation between their interest on demand deposits series and the one used here for the period 1950–68 is .89.

[8] In a reply to comments on his study, Lee (1969) enters r_s and r_M (i.e., service charges on demand de-

assume that the own price effect is greater in absolute value than the cross price effect on the demand for money, then our demand for money equation (10) implies that $\partial(M/P)^d/\partial i$ will be negative. An increase, for example, in the benchmark rate of return which increases both the price of monetary services from money and the price of monetary services from the substitute asset will then decrease the demand for money.

In the regressions that follow, I estimate equation (10) without making assumptions (a) and (b). The important unanswered question of whether commercial banks evade the current prohibition of interest payments on demand deposits is not begged and the economically meaningful P_M, P_s specification is tested.

II. Estimates of the Prices

I shall operationally measure the "pure" pecuniary marginal rate of return i by the yield on long-term (up to thirty-year) corporate bonds r_L. Our data requirements limit us to the use of highly marketable assets upon which long historical interest rates series exist. Therefore, although long-term corporate bonds represent what in commonly vague terminology is called the least "liquid" asset that fulfills this criterion, they can be assumed to yield significant nonpecuniary monetary service returns. Hence r_L underestimates the true "pure" pecuniary rate of return and is an essentially arbitrary and biased zero monetary service benchmark.[9] The marginal yield on a monetary substitute r_s will be operationally defined as the short-term (four-six month) commercial paper rate.[10] The rental price of monetary services from a dollar of monetary substitutes P_s is then equal to $(r_L - r_s)$.

The marginal return on money r_M is defined as a weighted average of interest on currency r_C and interest on deposits r_D.

$$r_M = \left(\frac{C}{M}\right) r_C + \left(\frac{D}{M}\right) r_D \tag{12}$$

The money supply M is defined as currency C plus commercial bank deposits D and is represented by M_1 if bank deposits are defined to include only demand deposits and by M_2 if bank deposits are defined to include demand and time deposits. Interest on currency is assumed to equal zero.[11] Interest on bank deposits is estimated by measuring the marginal costs of producing monetary services yielded by a dollar of deposits MC_D and assuming that perfect competition among banks forces each of them to pass on to their depositors, in an open or covert manner, all marginal profit from their deposit accounts; or equivalently, that the rental price of monetary services from a dollar of deposits is equal to the marginal costs of monetary services from a dollar of deposits.

$$(i - r_D) = MC_D \tag{13}$$

where r_D represents the "competitive" (or zero marginal profit) rate of interest paid on deposits by commercial banks.[12]

posits) as separate variables in demand for money regressions but still does not explicitly consider a benchmark return nor specify own and cross prices.

[9] The return on the much more illiquid asset human capital may be a superior measure of i, but it is extremely difficult to obtain reliable annual estimates of this return. If r_L is a return on a close substitute for money (i.e., a good alternative source of monetary services) rather than a measure of the benchmark rate of return, then $\partial(M/P)^d/\partial r_L$ rather than being indeterminate, will be negative. My particular empirical formulation therefore crucially depends upon how suitable r_L is as a measure of the benchmark rate of return.

[10] I experimented with alternative short-term interest rates in various subperiods and found the price of savings and loan shares to be statistically significant in the postwar period while the price of time deposits was found to be statistically insignificant in demand for M_1 regressions. Given the strong postwar trends in these series, however, any conclusions reached concerning substitutability solely on the basis of this evidence must clearly be tentative.

[11] Although some currency has at times paid interest. See Friedman and Anna Schwartz, p. 644, fn. 3.

[12] Individual commercial banks are assumed to be facing perfectly elastic demand curves. If competitive

The major element of MC_D is the interest foregone by commercial banks on the margin on the assets in their portfolio. Any asset which a bank holds that on the margin yields a rate of return less than i should properly be considered as a marginal cost of producing monetary services. Assume that a commercial bank's portfolio consists of two assets, non-interest bearing reserves (R) and investments (I) which yield a marginal rate of return of r_I, and that $R+I=D$. If (R/D) and (I/D) are the bank's marginal reserve to deposit and investment to deposit ratios, then the foregone interest cost per marginal dollar of deposit is $i(R/D)+(i-r_I)\ (I/D)$. If we assume that all other commercial bank costs are not marginal costs, then competitive interest payments on deposits are, from (13),[13]

$$r_D = r_I(1 - [R/D]) \tag{14}$$

If (R/D) is also equal to the average reserve to deposit ratio, then from (12)

$$r_{M_2} = r_I(1 - [H/M_2]) \tag{15}$$

where high-powered money (H) equals currency (C) plus total commercial bank reserves (R).

An estimate of r_{M_1} can be obtained from our general relationships, equations (12)–(14), by letting D represent solely demand deposits DD and replacing R with R_{DD}, reserves held by commercial banks against demand deposits.[14]

interest payments are not made on deposits then r_D would be less than $(i-MC_D)$ with the bank in disequilibrium earning a marginal rate of pure profit (i.e., marginal revenue is greater than marginal cost). Every firm in the banking industry would therefore want to expand output and new firms would want to enter the industry. We can assume that new entry is restricted by government licensing and that the market is shared among existing firms in nonprice ways (say, randomly). However, if banks are permitted to engage in nonprice competition, then the difference between price and marginal cost will be eliminated. As long as there were *some* type of variable expenditure by a bank which to the consumer was equivalent to an increase in r_D (however small), then each bank's attempt to expand its output would in the process raise MC_D (and to some extent r_D) and thereby eliminate any marginal profit. If individual commercial banks have rising marginal cost schedules, then total profit may remain positive. This represents a rent on the government franchise limiting entry. If individual commercial banks were assumed to face negatively sloped demand curves, then the profit-maximizing price of monetary services from deposits would remain greater than marginal cost. Each banking firm would have some type of specific capital (say, a particular location or brand name) which determines the position of its demand curve and therefore its market share. Some of the profit earned by the firm would then represent a rent on this specific capital and r_D would still be negatively related to MC_D. In operationally defining r_D changes over time in the degree of competitiveness of the banking industry are ignored.

[13] Alternatively, we can derive (14) by defining MC_D less broadly to equal the marginal reserve ratio multiplied by the marginal rate of interest that would be earned by commercial banks if they could replace the reserves with earning assets $r_I(R/D)$, and then by assuming that r_D will equal the difference between r_I and MC_D. This formulation has the drawback of misleadingly identifying commercial bank costs with reserves. But even if there were no reserve requirements and commercial banks held no reserves, the marginal cost of producing monetary services and the competitive rental price of monetary services from money would not be zero since r_I is generally less than i. Annual commercial bank operating expenditures (which, including taxes and dividends, have recently been about 5 percent of total outstanding deposits) are ignored in measuring MC_D because these expenditures are not generally marginal costs but are related to, for example, the number and type of transactions made. A perfectly competitive banking system would cover these expenditures by correctly pricing transactions. (If, for example, interest were paid via reductions in loan rates, highly active accounts would receive less of a reduction for the same average demand deposit balance than less active accounts). The marginal interest on deposits, defined in equation (14), is what should enter the demand for money, while an average interest on deposits variable, which would include the additional expenditures, would determine the number and type of deposit accounts (i.e., the all-or-none decision). In addition, some of these expenditures may represent for our calculations not costs of producing deposits but implicit interest payments on deposits in the form of "free service" to customers.

[14] Equation (16) would be equivalent in form to equation (15) and $(DD-R_{DD})/M_1$ could be written as $1-(H/M_1)$ if M_1 were defined consistently to include reserves held by commercial banks against time deposits as part of "currency held by the nonbank public" $(C=H-R_{DD})$. This definition would treat the high-powered money held by commercial banks against time deposits in a way logically equivalent to the present treatment of the cash holdings of nonbank financial intermediaries (for example, savings and loan institutions). However, I use the common, less inclusive but

$$(16) \qquad r_{M_1} = r_I[(DD - R_{DD})/M_1]$$

The marginal return on commercial bank investments r_I is measured by r_s because the *true* r_I is very difficult to obtain. Reported rates of return earned by commercial banks are inadequate because they are gross average rates of return and not a measure of the marginal profit a bank can earn. Such rates are also meaningless if reduced interest rates on loans is the major way in which interest on demand deposits is paid.

It should be emphasized that these estimates of the perfectly competitive marginal return on money assume that the regulations governing the maximum interest rates which banks are permitted to pay on their deposit liabilities, the prohibition of all interest payments on demand deposits, and the ceiling rate set by regulation Q on time deposits, are completely ineffective, i.e., are evaded costlessly. "Indirect" price competition by commercial banks is so efficient that marginal profit is eliminated entirely by increases in r_D rather than by increases in MC_D. But for these measures to be empirically valid in the regressions that follow, all that is necessary to assume is that the costs of making interest payments on deposits have not changed significantly over time.

The returns on money defined in equations (15) and (16) also assume that commercial bank reserves yield no interest. If however, the Federal Reserve subsidizes member commercial banks in proportion to their reserves, then this subsidy will be passed on to depositors by the same competitive process by which we assume interest is being paid on deposits. Such subsidy payments must therefore be subtracted from MC_D and the r_D estimate revised upward. Our r_D estimates are therefore modified for the subsidies given by the Fed to member commercial banks in the form of credit extended at less than the market interest rate. These include bank float which is extended at a zero interest rate and borrowings by commercial banks at the discount rate.[15] Modifications are also

possibly more economically meaningful, definitions of C and M. (Note that any attempt to estimate the demand for the more inclusive M_1 magnitude would have to include legal reserve requirements on commercial bank time deposits as an explanatory variable.) Similarly, M_2 is defined exclusive of large negotiable certificates of deposit for the 1961–70 period and therefore, during this period, R is defined as consistent with the deposit total in our monetary aggregate and does not include required reserves held on large *CD*s. These reserves are also not considered as part of C and hence are totally excluded from H. To estimate R_{DD} one must allocate excess commercial bank reserves between demand and time deposits. I construct R_{DD}, and therefore r_{M_1}, by allocating excess reserves between the different deposits in proportion to the relative quantities of the deposits. This may be considered to yield a somewhat low estimate of excess reserves held against demand deposits. Two alternative but somewhat more arbitrary estimates were also made: (a) allocating excess reserves in proportion to required reserves, and (b) allocating all excess reserves to demand deposits. Demand for money regressions comparing the three alternative estimates yield almost identical results except for the period from 1933 to 1942 when excess reserves were significant. And for that period the estimate we have used in constructing r_{M_1} yields slightly superior results.

[15] This credit has at times been significant. Through the 1960's float averaged more than 10 percent of total member bank reserves, and for the first few years after the Federal Reserve was established, borrowings were greater than reserves. This latter phenomenon is analogous to the usual "discounts" new firms often make to obtain a market share. Borrowings were greatest from 1918 to 1921, while the Federal Reserve's share of the deposit market was rising from 45 to 65 percent, and taking account of borrowings helps to explain some of the increase in the demand for money in this period. Float is subtracted from member bank reserves by the Fed. This definition of bank reserves is somewhat misleading. Bank float should be considered as a part of member bank reserves and subtracted out only when making calculations of commercial bank costs. Regressions which include float in reserves, however, yield slightly superior results than the reported regressions which exclude float. The hypothesis that float is an interest-free loan of reserves to member banks from the Fed is therefore rejected. But these results are dominated by the post-1950 period when float becomes significant and exclusion produces poorer results by exacerbating the unexplained postwar decrease in the demand for money. In my 1972 paper, I show that one possible partial explanation for the postwar increase in velocity

made for the subsidy from the Treasury in the form of federal government demand deposits held at commercial banks. If we assume that the federal government does not participate in the evasion of interest prohibition on demand deposits, then the interest earnings from government demand deposits, adjusted for the reserves held against them, are a subsidy to commercial banks.[16]

Finally, estimates of the marginal return on money are further modified by subtracting the service charges that individuals pay on demand deposits and the losses that they expect to incur in holding deposits if the bank fails. These elements can be considered as covering additional commercial bank marginal costs of producing deposits.[17]

III. Results

Estimates of equation (10) using annually averaged data and the definitions outlined above are presented in Table 1. The time period 1880–1970 is covered for M_2, with subperiods 1880–1918 for M_2 and 1919–70 for M_1 and M_2 reported separately. (The strong unchanging trends in the variables over shorter periods cast serious doubt on the effective number of degrees of freedom present within smaller subperiods.) Money and income are defined in real permanent per capita terms; i.e., the demand function is assumed homogenous of the first degree in permanent prices and measured population.[18]

may be a decrease in the "quality" of money (measured by the predictability of prices).

[16] These deposits averaged \$6.4 billion in 1970, or more than 20 percent of total commercial bank reserves. The largest banks generally hold the largest quantities of government deposits and the deposits can be considered as a subsidy given approximately in proportion to reserves held. The large increase of government demand deposits during 1943–46 (when the ratio of government demand deposits to total bank reserves reached a level of nearly one) and the subsequent sharp postwar decrease in these deposits helps to explain movements in the demand for money during this period. The other relatively large increase in government demand deposits occurred in 1918 (when the ratio of government demand deposits to total reserves rose to more than one-half), and, although real cash balances decline sharply, inclusion of this variable improves the fit.

[17] There are, however, a priori reasons to suggest that service charges should not be subtracted out. Service charges face the depositor in per transaction or per unit time and not per dollar deposit terms and therefore do not enter the relevant marginal P_M. If the number of transactions an individual makes per unit time and the average balance he holds are substitutes, the effect on the demand for money of service charges should be in the opposite direction. A rising service charge rate per transaction would decrease the number of transactions and increase the average deposit balance. This may possibly be tested by looking at total bank clearings and average level of demand deposits as service charges change. However, in my work regressions which do not subtract service charges from estimates of r_D yield poorer results than those which do. Service charges monotonically increased from .15 percent of demand deposits in 1946 to .61 percent of demand deposits in 1967, where they have stabilized. Hence inclusion helps to "explain" some of the postwar rise in velocity. Service charges on deposits may therefore actually enter the demand for money as negative interest payments (or as a reduction in "free" services) as Selden, Lee, Feige (1964), and Phillip Cagan consider them to be. But, as we shall see below, an "actual" interest payment on money variable does not work very well over the 1919–70 period. And when alternative explanations for the changing postwar trend in velocity are considered (see the author (1972)), service charges lose their explanatory power.

[18] Changes in permanent prices and measured population are assumed to cause equal percentage changes in the nominal quantity of money demanded. Some evidence on this proposition with respect to measured prices can be found in Allan Meltzer. Friedman (1959) argues that holders of cash balances are likely to determine the quantity to hold in terms of longer term price movements, and successfully uses a permanent price index deflator. But Friedman's series, which is merely a weighted average of past measured price levels with no adjustment made for trend, cannot be meaningfully extended through the 1960's when a dollar fiduciary standard with a clear positive long-term trend in the price level replaced any remaining semblance of a gold commodity standard with its general presumption of a long-term stable price level (see the author (1974)). (The downward bias is obvious when one considers the level of the ratio of measured to unadjusted permanent prices, which is greater than one every year since 1940, is 1.03 in 1963 before the rapid inflation begins, and reaches a level of 1.08 in 1970.) Any attempt to adjust the permanent price level series must be somewhat arbitrary. I crudely use the accelerating level of the long-term interest rate as a market measure of the postwar acceleration of the long-term trend of prices and measure trend over the 1952–70 period by subtracting the long-term rate of interest from the average long-

TABLE 1—DEMAND FOR MONEY FUNCTIONS, ANNUAL OBSERVATIONS, 1880–1970 AND SUBPERIODS

Time period	$\log M = a_0 + a_1 \log y_p + a_2 P_s + a_3 P_M$ a_0	a_1	a_2	a_3	R^2	$D.W.$	Standard Error
M_2							
1880–1970	−13.92	1.328	.327	−.342	.988	1.05	.0773
[.6957]	(78.90)	(52.98)	(16.67)	(14.01)			
1880–1918	−15.20	1.515	.214	−.277	.987	1.15	.0523
[.4413]	(21.21)	(13.83)	(6.09)	(6.18)			
1919–70	−13.82	1.312	.353	−.350	.954	.70	.0762
[.3432]	(49.18)	(31.29)	(8.86)	(7.08)			
M_1							
1919–70	−15.68	1.562	.416	−.446	.905	.50	.1254
[.3956]	(32.59)	(21.06)	(11.83)	(9.25)			

Note: All *logs* stand for natural logarithms. The absolute values of the *t*-statistics are given in parentheses beneath the coefficient estimates; *D.W.* represents the Durbin-Watson statistic; R^2 the coefficient of determination. The standard error of the dependent variable for each time period is bracketed beneath each date.

The regressions are not run with the *logs* of P_s and P_M because these variables are sometimes negative. This points up the fact that r_L is an arbitrary base from which to measure monetary services. But some zero benchmark must be chosen, and as long as the monetary service yield in the benchmark asset remains constant over time, the only effect on P_M and P_s will be a once and for all decrease in their levels. Elasticity expressions are, however, not invariant under changes of origin and therefore this is one reason for running demand for money regressions in semilog form.[19]

term rate during 1946–51, smoothing a monotonically increasing series to the nearest quarter of a percentage point. The adjusted permanent price level series remains below measured prices, but the measured to permanent price ratio is only 1.01 in 1963 and rises to only 1.04 in 1970. Surprisingly, the long-term interest elasticity of demand for money is not increased very much over this period when this is done. The demand for real per capita M_1 using Friedman's unadjusted permanent prices for 1953–70 is:

$$\log M_1 = \underset{(3.51)}{-13.82} + \underset{(2.22)}{1.319} \log y_p - \underset{(3.62)}{1.474} r_s - \underset{(1.60)}{.045} r_L + \underset{(3.72)}{2.128} r_{M_1} \qquad R^2 = .661$$

while using my adjusted permanent prices it is:

$$\log M_1 = \underset{(3.39)}{-13.66} + \underset{(2.13)}{1.300} \log y_p - \underset{(3.62)}{1.509} r_s - \underset{(1.74)}{.050} r_L + \underset{(3.71)}{2.179} r_{M_1} \qquad R^2 = .701$$

With regard to the deflation by measured population, the use of per capita averages of aggregated income and money holdings may not be an appropriate approximation. Some measure of the dispersion of income and money holdings may also be relevant. In addition, since foreign holdings of currency are included in the money estimates, the *U.S.* population figures used are somewhat arbitrary.

[19] "Since there are no natural zeros from which we measure economic magnitudes, the elasticity expressions can be seen to be essentially arbitrary" (Paul Samuelson, p. 125). Modern portfolio theory does not suggest that interest rates or interest differentials should enter in a logarithmic form and there is some economic reason to expect a one percentage point change in P_M, independent of the level at which it occurs, to represent the same increase in the cost of holding money and to have the same effect on the demand for money. The commonly used logarithmic functional form implies a proportionately greater effect for every percentage point change in interest the lower the rate of interest and an undefined demand for money at a zero rate of interest. But as long as there are some increasing marginal costs associated with holding cash balances, we should expect the demand for money curve to cut the axis and a finite determinant money demand at negative interest rates. In addition, once we introduce risk considerations associated with the future real value of money, even at a zero (or negative) net cost of holding money, an optimum portfolio will not consist entirely of money. If there is a negative covariance between the movements of the price level and the return on physical

TABLE 2—COMPARISON OF TABLE 1 DEMAND FOR MONEY RESULTS WITH BENCHMARK DEMAND FOR MONEY RESULTS

Dependent Variable and Time Period	Table 1 y_p	Table 1 P_s	Table 1 P_M	Benchmark y_p	Benchmark r_s	Table 1 $S.E.$ / Benchmark $S.E.$
M_2						
1880–1970	1.33	.33	−.34	1.52	−.06	.64
	(53.0)	(16.7)	(14.0)	(47.8)	(8.4)	
1880–1918	1.52	.21	−.28	2.13	−.03	.76
	(13.8)	(6.1)	(6.2)	(38.3)	(2.5)	
1919–70	1.31	.35	−.35	1.27	−.06	.94
	(31.3)	(8.9)	(7.1)	(29.2)	(10.1)	
M_1						
1919–70	1.56	.42	−.45	1.31	−.10	.84
	(21.1)	(11.8)	(9.3)	(16.5)	(9.7)	

All of the estimates of the income elasticity and the logarithmic price slopes of demand for money in Table 1 have the correct sign and are statistically significant at the .99 confidence level. But to interpret these results they must be compared to some standard estimates of the demand for money. As a first step I will use estimates of equation (11) as a benchmark, with *the* interest rate entering in a semilog function. Table 2 compares these benchmark estimates with the estimates of Table 1. At the end of each row is the ratio of the standard error of estimate of the regression reported in Table 1 to the standard error estimate of the benchmark regression. This number can be thought of as a measure of the reduction in the sum of squares achieved by using the demand for money function we have been estimating, compensating for the sacrifice of a degree of freedom. If it is less than one, it means that the results reported in Table 1 have more explanatory power than the benchmark results. The results reported in Table 1 improve upon the benchmark results in all periods, with the regression over the entire 1880–1970 time period showing the most dramatic improvement. This is possibly the most meaningful regression, since much of the variation that must be explained occurs between the subperiods.[20]

The demand for M_2 estimates of Table 1 also show substantially more stability over time than the benchmark estimates, with

assets, individuals who dislike risk will diversify and hold a fraction of their wealth in common stocks and other real assets, even if the return is lower. The Keynesian analysis often cited to justify a logarithmic demand for money function considers nominally denominated bonds as the alternative to money and ignores the risk associated with unanticipated price level changes. Riskiness consists solely of the capital losses associated with interest rate changes and money is assumed to be a riskless asset. The *log-log* interest rate results were highly erratic and generally much poorer than the results in Table 1. For example, over the 1880–1970 time period:

$$\log M_2 = \underset{(63.60)}{-14.30} + \underset{(44.33)}{1.419} \log y_p - \underset{(7.19)}{.724} \log r_s$$

$$- \underset{(1.21)}{.088} \log r_L + \underset{(6.10)}{.481} \log r_{M_2}$$

$$R^2 = .983;\ DW = .50;\ SE = .0930$$

[20] Although first differencing may eliminate much of the information (I consider my r_M variable to be a proxy for the general level of interest payments on money and have much less confidence in its year to year movements), the first difference results over the 1880–1970 period are:

$$\Delta \log M_2 = \underset{(1.92)}{.009} + \underset{(7.33)}{1.030} \Delta \log y_P + \underset{(4.25)}{.080} \Delta P_s$$

$$- \underset{(4.03)}{.099} \Delta P_{M_2}$$

the F-statistic for identity of coefficients between the two subperiods falling to 8.1 from 45.8 ($F_{.01}=3.6$). Much of this improvement is due to the decrease in the estimated income elasticity for 1880–1918, which is produced by the sharp decline that occurred in H/M_2 over this period.[21] The estimated income elasticity for 1880–1918, however, remains greater than the income elasticity estimate for 1919–70. This is expected since money like all goods should have a declining income elasticity of demand as income rises; entering the consumer's budget as a luxury, but eventually becoming a necessity.[22] It is therefore misleading to ask what *the* income elasticity of demand for money is.

$R^2=.514$; $DW=1.34$; $SE=.0358$; SE/Benchmark $SE=.94$. This regression is presented so that it can be compared to previously reported results and not to correct for serial correlation. The very low $D.W.$ statistic indicates a high degree of serial correlation in the errors and therefore our OLS estimates do not have minimum variance. But although the precision of our coefficient estimates is less than indicated by the reported standard errors, it should be noted that the reported T-values are extremely high. The GLS estimates, using the Cochrane-Orcutt iterative technique, for the 1880–1970 period are:

$$\log M_2 = -\underset{(16.90)}{12.330} + \underset{(9.97)}{1.063}\log y_P + \underset{(3.89)}{.072}P_s - \underset{(3.65)}{.087}P_{M_2}$$

$R^2=.998$; $DW=1.37$; $SE=.0343$; $\rho^*=.96$

$$\log M_2 = -\underset{(17.86)}{12.634} + \underset{(10.89)}{1.137}\log y_P - \underset{(2.33)}{.044}r_s - \underset{(4.37)}{.056}r_L + \underset{(2.12)}{.052}r_{M_2}$$

$R^2=.998$; $DW=1.30$; $SE=.0320$; $\rho^*=.96$

[21] See the author (1973) where it is also demonstrated that once H/M_2 is included in a velocity relationship the seemingly inconsistent movements in velocity and the rate of interest emphasized by Friedman and Schwartz are reconciled. The H estimates used in this earlier work mistakenly included bank float for 1968–70 and thereby exacerbated the unexplained discrepancy that remains between velocity and interest rate movements over this recent period.

[22] If income elasticities of luxury goods do not fall, then as income grows we should see individuals specializing in the consumption of these goods. But since the weighted sum of the income elasticity of demand for all goods must equal one, in the limit as the luxury goods weights increase, the income elasticities must fall to one. This may partially explain the postwar rise in velocity. The downward trend in the income parameter should produce some additional serial correlation in the residuals in the regressions over the entire period since these regressions use an average income elasticity which underestimates the true income elasticity in the early years and overestimates the true income elasticity in the later years.

The regressions reported in Table 1 are simultaneously testing both hypotheses: the significance of the competitive interest return on money variable and the functional form specification embodying the price of monetary services. I shall now attempt to test each hypothesis separately by running the same regressions in the more general semilogarithmic interest rate form with r_L, r_s, and r_M, in addition to $\log y_P$, entering as separate variables. The results are reported in Table 3. Although the interest on money variable remains significant in all but one of the time periods (1880–1918), the interest rate functional form appears to outperform the price functional form with a lower standard error of estimate for all four regressions. In addition, the estimated demand for M_2 becomes slightly more stable over the two subperiods with the F-statistic falling to 7.0 ($F_{.01}=3.3$). However, the Durbin-Watson statistic is now lower for all four regressions and it is therefore not unambiguously clear which functional specification is superior.

The much lower magnitude of the r_L coefficient than of the r_s or r_M coefficients for the demand for money regression over the entire 1880–1970 period suggests that it may be appropriate to consider long-term corporate bonds as we have, i.e., as a benchmark asset yielding no monetary services rather than as a close substitute for money. The closeness of the r_L co-

TABLE 3—DEMAND FOR MONEY FUNCTIONS, ANNUAL OBSERVATIONS, 1880–1970 AND SUBPERIODS

	$\log M = a_0 + a_1 \log y_p + a_2 r_s + a_3 r_L + a_4 r_M$							
Time Period	a_0	a_1	a_2	a_3	a_4	R^2	$D.W.$	Standard Error
M_2								
1880–1970	−14.09	1.372	−.285	−.058	.303	.990	.94	.0721
	(82.50)	(52.44)	(13.28)	(4.06)	(12.09)			
1880–1918	−16.87	1.829	−.110	−.130	.131	.991	.74	.0451
	(21.85)	(14.23)	(2.62)	(5.73)	(2.33)			
1919–70	−13.59	1.287	−.208	−.039	.200	.960	.46	.0720
	(48.56)	(31.56)	(3.11)	(1.88)	(2.69)			
M_1								
1919–70	−14.31	1.381	−.165	−.129	.170	.918	.32	.1181
	(20.87)	(14.23)	(1.66)	(3.02)	(1.51)			

efficient to zero also indicates another alternative functional form of the demand for money equation that should be tested. Although the magnitude of the P_M coefficient is generally greater in absolute value than the P_s coefficient (as one would expect), it is not much greater. The sum of a_2 and a_3 is negative in three out of four regressions, but is never greater in absolute value than 0.07. Clearly the two price effects offset one another to a large extent. And it is possible that although each coefficient separately has a high t-value, the sum which is numerically much smaller might not be statistically significant at all. The significance of a_2 and a_3 separately might merely reflect the fact that r_s enters both prices, in one case positively affecting the demand for money and in one case negatively affecting the demand for money. I therefore tested this possibility by running regressions using the difference between the price of money and the price of money substitutes $(P_M - P_s)$ in addition to income as the relevant independent variable. Since $(P_M - P_s)$ is equivalent to $(r_s - r_M)$ I am also testing the Selden-Lee assumption that the relevant price variable is the difference between the rate of return on a substitute asset and the rate of return on money. If the t-ratio of $(a_3 + a_4)$ were less than one then a regression using only $(r_s - r_M)$ would give better results than the more complete and complicated regressions of Table 1.[23] The results are reported in Table 4. They are generally inferior to those in Table 1 with higher standard errors of estimate for three of four regressions, implying that one should, in general, "pay" the extra degree of freedom to obtain the separate price effects on the demand for money. But the Selden-Lee hypothesis that $a_2 = -a_3$ cannot be rejected using an F-test at the 99 percent confidence level in three of four regressions. The empirical suitability of the Selden-Lee functional form specification (not the particular r_M variables they use) remains an open question.

Finally, to test if the maximum interest payment regulations on deposits are being evaded, we can compare our results for the most recent time period with results obtained assuming that the regulations are effective and that no implicit interest payments are being made by banks. The demand for money over 1919–70 using actual reported interest payments on deposits, modified by expected losses and service charges in constructing an actual interest payments on money variable (r_{MA}) is:

[23] Note that $a_2 P_s + a_3 P_M = (a_2 + a_3) P_s + a_3 (P_M - P_s)$. See Yoel Haitovsky.

TABLE 4—DEMAND FOR MONEY FUNCTIONS, ANNUAL OBSERVATIONS, 1880–1970 AND SUBPERIODS

Time Period	$\log M = a_0 + a_1 \log y_P + a_2 (r_s - r_M)$ a_0	a_1	a_2	R^2	D.W.	Standard Error
M_2						
1880–1970	−13.96	1.328	−.315	.988	1.12	.0780
	(78.81)	(52.48)	(17.05)			
1880–1918	−17.17	1.815	−.139	.981	.70	.0630
	(26.45)	(18.28)	(3.83)			
1919–70	−13.81	1.312	−.356	.954	.69	.0755
	(50.34)	(31.61)	(11.11)			
M_1						
1919–70	−15.66	1.547	−.399	.902	.54	.1263
	(32.33)	(20.95)	(12.09)			

$$(17) \quad \log M_2 = -\underset{(37.43)}{13.19} + \underset{(22.84)}{1.220} \log y_p - \underset{(1.50)}{.022} r_s - \underset{(2.06)}{.046} r_L - \underset{(1.04)}{.038} r_{M2A}$$

$$R^2 = .954;\ DW = .29;\ SE = .0764$$

A comparison of these regressions with the third line in Table 3 demonstrates the superiority of the competitive interest on money variable to the actual interest on money variable.[24] This suggests the conclusion that the complete lack of effectiveness of interest rate regulations is closer to the truth than the commonly accepted assumption that it is a totally effective constraint on bank behavior.[25] But, more generally, the two alternative hypotheses we compared can be thought of as representing two extremes on a continuum; and some third hypothesis in between the costless evasion and no evasion hypotheses, may work best of all. Direct estimates of the interest paid on deposits, say, via reduced loan rates, would therefore be extremely useful.

IV. Summary and Implications

The significance of the "competitive" r_M variable in the demand for money function, whether entered indirectly as in Tables 1 and 4 or directly as in Table 3, is the primary conclusion of this paper. This represents a substantial advance, since no alternative interest on money series covering as long a time period now exists. The series presented here (and listed in the Appendix) is shown to make a significant improvement over the common assumption that interest on demand deposits equals zero, or the negative of service charges and expected losses on deposits.

My particular demand for money formulation also clarifies, to some extent, the debate which has occurred on whether a short-term or a long-term interest rate is the "proper" variable to use in the demand for money.[26] Most of the discussion has

[24] A measure of actual interest payments on money not modified by service charges and expected losses yielded even poorer results. Demand for M_1 regressions using an actual interest payments on money variable cannot be run over the 1919–70 time period because interest payments on demand deposits were not reported separately from interest payments on time deposits until 1927. Actual interest payments on M_1, however, did perform poorly over the 1927–70 period. A composite interest payments on M_2 variable, defined as a weighted average of actual reported interest payments on time deposits and my perfectly competitive interest payments on demand deposits, performed poorest of all over this period.

[25] Since $(1-[H/M])$ is merely a proxy for r_M/r_s, r_M could actually equal, for example, $(.01)\ r_s(1-[H/M])$ and yield significant regression results although the interest rate regulation is almost completely effective. This, however, seems highly unlikely given the fact that the magnitude of the P_M and P_s coefficients are nearly the same.

[26] For example, Martin Bronfenbrenner and Thomas Mayer and Ronald Teigen use a short-term interest

been almost completely devoid of any theoretical argument and the issue has generally been settled in favor of a short-term rate on empirical grounds.[27] The model I have presented suggests a theoretical reason for including both interest rates in the demand for money; a long-term interest rate as a measure of the rate of return on an asset yielding no monetary services and therefore as the opportunity cost of holding an asset yielding monetary services, and a short-term interest rate as a measure of the rate of return on an asset yielding a significant monetary service flow and therefore as the foregone pecuniary return on close substitutes for money.

In addition, my demand for money formulation points up the deficiencies in the usual interpretations of what economists call the interest elasticity of demand for money. Estimates of such a concept are obtained from regressions which do not consider the rental prices of the monetary service stream yielded by assets, but merely the rates of interest paid on assets. And such regressions generally omit from consideration the interest paid on money; therefore measuring a cross rate effect without holding the own rate constant. Because interest payments on money and on money substitutes are highly positively correlated with one another and influence the demand for money in opposite directions, such "observed" interest elasticities will be severely biased downward.[28]

But recognition of the fact that the *true* interest elasticity of demand for money, estimated while holding interest payments on money constant, is much greater than what has been estimated by previous studies should not necessarily be considered as a refutation of those who believe *the* interest elasticity of demand for money to be small. These results emphasize that the magnitude of the effect will depend upon what one considers to be held constant when carrying out an experiment. The usefulness of a particular specification of *ceteris paribus* conditions depends upon the purpose the calculation is to be used for. For policy considerations it may be the actual observed interest elasticity which is relevant. Policy makers may only wish to know if changes in the level of interest rates will cause (i.e., be associated with) large changes in velocity. My model supplies a theoretical justification for expecting a small total measured interest rate effect on the demand for money and the results supply an empirical justification for the net effect remaining negative; i.e., for the interest rate on money variable not dominating. An increase, for example, in the anticipated inflation rate unambiguously produces a decrease in the demand for money when the demand function is specified in the price form of Table 1 as long as the P_M coefficient is negative. Since

rate while Meltzer, Karl Brunner and Meltzer, Robert Eisner, Carl Christ and Henry Latané use a long-term rate.

[27] See, for example, Laidler (1966b), H. R. Heller, and Lee (1967).

[28] Since we use r_s as a measure of r_I, the marginal rate of return earned by commercial banks on their investment portfolio, r_s and r_M must *necessarily* be positively correlated. An alternative measure of r_I would make the proposition an empirical one. To get an idea of the bias, consider, for example, the following stepwise results for the 1919–70 period.

$$\log M_1 = \underset{(26.70)}{-\,14.21} + \underset{(16.50)}{1.313}\log y_p - \underset{(9.69)}{.100}\, r_s$$

$$R^2 = .866;\ DW = .35;\ SE = .1477$$

$$\log M_1 = \underset{(27.38)}{-\,15.63} + \underset{(17.57)}{1.542}\log y_p - \underset{(5.60)}{.392}\, r_s + \underset{(4.20)}{.390}\, r_{M_1}$$

$$R^2 = .902;\ DW = .54;\ SE = .1276$$

Once we take account of interest payments on money, the short-term interest rate effect on the demand for money is multiplied in magnitude by a factor of nearly four. Note that since the regression is in semilog form, the coefficient on the short-term interest rate does not represent the interest elasticity, but rather an estimate of the logarithmic slope; i.e., a measure of the percentage change in the quantity of money demanded from a percentage *point* change in the interest rate. If $\log M = br$, then $d \log M/dr = b$. Since $d \log r = dr/r$, the interest elasticity $d \log M/d \log r$ then equals br. Evaluated at the sample means over the entire 1880–1970 period the respective interest elasticities are -1.086 for r_s, $-.224$ for r_L, and $+.821$ for r_{M_2}.

r_L and r_s will increase by the same amount (when interest rates are stated in continuous terms), P_s will be unchanged while r_M will increase by less than r_L (at the sample mean r_{M_2} will rise by the increase in the expected inflation rate multiplied by $\bar{r}_{M_2}/\bar{r}_s$ or .711, which is $(1-H/M_2)$ in the simplest case), and P_M will therefore rise. Our price results imply that a one percentage point increase in the expected inflation rate will increase the demand for money .099 percent; i.e., (.342)(.289). In terms of the interest rate specification of Table 3, there is no such obvious negative relationship but our results imply that, again evaluated at the sample means for the 1880–1970 period, a one percentage point increase in the expected inflation rate will decrease the demand for money .128 percent; i.e., $.285+.058-(.303)(.711)$.[29]

This competitive banking model also implies that policy makers must be concerned with the effects of particular policy changes on commercial bank costs. Changes in bank costs will influence interest payments on money and therefore the demand for real cash balances. An increase, for example, in reserve requirements offset by open market purchases which keeps the money supply constant will raise the costs to banks of producing deposits and by leading to a fall in the rate of interest paid on deposits and therefore in the demand for money will have an expansionary effect on the economy. That is, changes in the composition of money will have nonneutral effects. And, in general, one must know how a particular change in the money supply is brought about to fully determine its effects.[30]

My results also have implications for the definition of money. The proper definition of money has become an essentially empirical issue of finding that particular subgroup of assets which yields a relatively simple and stable demand function. And current debate has narrowed down to the question of whether to include commercial bank time deposits as part of the money supply. Past results on this question have been somewhat ambiguous. For example, Laidler concludes that "... the stability of the demand function for money is improved by including time deposits in the definition of money" (1966a, p. 55) while Brunner and Meltzer argue that "... our results seem to suggest clearly that currency plus demand deposits is the more appropriate definition" (p. 390).[31] Brunner and Meltzer, however, also argue on theoretical grounds for the narrower definition of money. They say that the inclusion of interest yielding time deposits obscures part of the substitution effect between money and nonmoney, i.e., to "... mix the effects of general and relative changes in interest rates and to obscure a part of the wealth adjustment process" (p. 350;

[29] These calculations assume that H/M will be invariant to changes in the expected inflation rate. But since interest on currency is assumed to equal zero, changes in the rate of expected inflation will decrease the relative price of deposits compared to currency, thereby producing a decrease in the desired currency to deposit ratio and a decrease in H/M. The increase in r_M at the sample mean will therefore be somewhat greater and the net effect on the demand for money therefore be somewhat less than indicated by our calculations.

[30] Note that if the same experiment is analyzed in terms of a Patinkin-type model, the implications are reversed. Changes in the "inside" to "outside" (or interest bearing to non-interest bearing) money ratio disturb neutrality via a positive wealth effect on the demand for money. Also note that changes in the discount rate will have some economic impact in addition to any "psychological" or "announcement" effects. An increase in the discount rate increases commercial bank costs (reduces the subsidy from the Fed), raises the competitive price of money and is therefore, *ceteris paribus*, expansionary.

[31] Some of the disagreement is due to alternative definitions of the other variables in the demand function—a short-term or a long-term interest rate, nonhuman wealth or permanent income. Laidler attempts to compare alternative formulations and concludes that permanent income is superior to nonhuman wealth (1966a) and then using permanent income finds a short-term interest rate and a definition of money that includes time deposits to yield the most stable relationship (1966b).

also see Meltzer, p. 225). This is analogous to what I have identified as confusing the own and cross effects of a change in interest rates and a reason why observed interest elasticities of demand for money are underestimates of "true" interest elasticity. But Brunner and Meltzer's argument justifies a definition of money which excludes time deposits only if interest payments are not also made on demand deposits. This is an empirical hypothesis that is not true pre-1933 and which has been tested for the post-1933 period and found to be unwarranted. If interest is paid on demand deposits then we have the same problem of "obscuring the substitution effect" with respect to demand deposits; and Brunner and Meltzer are left with currency as the sole legitimate component of the money supply. A much better method to isolate a pure interest rate effect is not to eliminate from consideration as money all assets which yield interest but to explicitly introduce the own price of money. Once this is done, no particular definition of money is implied.

All single equation estimates of the demand for money, such as those reported here, run the risk of the "identification problem." I estimate the demand for real cash balances using annually averaged data and therefore am assuming that most, if not all, of the market adjustment takes place within a year, and that individuals are *on* their demand curves. My regressions are therefore estimates of a long-run or "equilibrium" demand for money.[32] But since H/M, the inverse of the money multiplier, now enters on the demand side as well as the supply side, the identification problem is magnified. Increases in R/D or C/D which will cause decreases in the supply of money will now also cause decreases by a competitive banking system in interest payments on money and hence a decrease in the demand for money. For example, if H is taken as exogenous, then the nominal supply of money and hence H/M is partially determined by the holders of money along with real cash balances. A more complete simultaneous model would therefore include a demand for money function and a demand for currency relative to deposits function, which partially determines H/M and which therefore feeds back into the demand for money. If a desired commercial bank reserve ratio is introduced, it will partially determine bank costs and hence interest payments on deposits, and therefore will feed into the demand for money and the demand for currency relative to deposits functions. The competitive banking model we are using here emphasizes the necessity of developing a general equilibrium model of the monetary system.

[32] Feige (1967) suggests that "... cash balance portfolio adjustments to desired positions are completed within a single year" (p. 471). Also see Lee (1967, p. 1177).

Appendix

Construction of the Interest Rate Paid on Money Variables

Competitive Interest Paid on M_2

$$\begin{aligned} r_{M_2} = {} & r_s[1-(H_2/M_2)] + (r_s^w - r_d^w)(B/M_2) \\ & + r_s[(G_{DD}-R_G)/M_2] \\ & - l_D^*(D/M_2) - s_{DD}(DD/M_2) \end{aligned}$$

Competitive Interest Paid on M_1

$$\begin{aligned} r_{M_1} = {} & r_s(DD-R_{DD})/M_1 \\ & + (r_s^w - r_d^w)(B/M_1)(R_{DD}/R_D) \\ & + r_s(G_{DD}-R_G)(1/M_1)(R_{DD}/R_D) \\ & - l_D^*(DD/M_1) - s_{DD}(DD/M_1) \end{aligned}$$

where:

$M_2=C+D$, currency plus commercial bank demand and time deposits, less large negotiable certificates of deposit beginning in 1961 (Friedman and Schwartz, *Fed. Res. Bull.*).

$M_1=C+DD$, currency plus commercial

bank demand deposits (Friedman and Schwartz, *Fed. Res. Bull.*).

r_s = rate of interest on short-term (four–six month New York City) commercial paper (Friedman and Schwartz).

$H_2 = C + R_D$, currency plus reserves held by commercial banks on deposits; where R_D equals vault cash plus deposits at Federal Reserve banks, excluding commercial bank float and reserves held on large certificates of deposit, assuming since 1966 that all large CD's are issued by banks with more than $5 million of deposits (Friedman and Schwartz, *Fed. Res. Bull.*).

R_{DD} = reserves held by banks on demand deposits; where member bank excess reserves and nonmember bank vault cash are allocated between demand and time deposits in proportion to the relative quantities of demand and time deposits, and member bank required reserves are estimated by subtracting the product of member bank time deposits and the effective required reserve ratio on time deposits from total member bank required reserves (Friedman and Schwartz, *Banking and Monetary Statistics*, *Fed. Res. Bull.*, *FRB* Surveys of "Time and Savings Deposits," *FDIC Assets and Liabilities of Commercial and Mutual Savings Banks*).

B = member bank borrowings at Federal Reserve banks (*Banking and Monetary Statistics*, *Fed. Res. Bull.*).

r_d^w = weighted annual average of discount rate at Federal Reserve Bank of New York, where weights are determined by monthly borrowings in the System (*Banking and Monetary Statistics*, *Fed. Res. Bull.*).

r_s^w = weighted annual average of monthly short-term commercial paper rate, where weights are determined by monthly borrowings in the System (Friedman and Schwartz, *Banking and Monetary Statistics*, *Fed. Res. Bull.*).

G_{DD} = government demand deposits at commercial banks, adjusted for a 1 percent interest payment by banks, 1909–32 (Friedman and Schwartz, *Fed. Res. Bull.*).

R_G = reserves held by commercial banks against government demand deposits (assumed to be the same as the average for ordinary demand deposits, i.e., $R_G = (R_{DD}/DD)(G_{DD})$, except for the particular years when *U.S.* government demand deposits were exempted from reserve requirements and $R_G = 0$).

Competitive Interest Payments on Money 1880–1970

Date	r_{M2}	Date	r_{M2}
1880	2.73	1900	3.05
1881	2.88	1901	3.08
1882	3.14	1902	3.64
1883	3.17	1903	4.07
1884	2.89	1904	3.10
1885	2.24	1905	3.32
1886	2.87	1906	4.32
1887	3.51	1907	4.82
1888	2.94	1908	3.22
1889	3.01	1909	2.98
1890	3.56	1910	3.77
1891	3.37	1911	3.06
1892	2.51	1912	3.66
1893	4.37	1913	4.34
1894	1.80	1914	4.35
1895	2.18	1915	2.71
1896	3.74	1916	2.70
1897	2.15	1917	3.85
1898	2.45	1918	4.89
1899	2.86		

Date	r_{M2}	r_{M1}	Date	r_{M2}	r_{M1}
1919	4.48	4.20	1945	0.51	0.45
1920	5.95	5.53	1946	0.55	0.48
1921	5.25	4.84	1947	0.63	0.54
1922	3.56	3.29	1948	0.90	0.77
1923	4.00	3.68	1949	0.93	0.79
1924	3.11	2.85	1950	0.92	0.79
1925	3.24	2.98	1951	1.42	1.26
1926	3.44	3.16	1952	1.56	1.38
1927	3.27	2.99	1953	1.70	1.49
1928	4.02	3.67	1954	1.01	0.85
1929	4.82	4.42	1955	1.48	1.28
1930	2.80	2.56	1956	2.32	2.04
1931	1.79	1.59	1957	2.69	2.35
1932	1.77	1.54	1958	1.68	1.41
1933	0.52	0.39	1959	2.88	2.49
1934	0.34	0.25	1960	2.79	2.38
1935	0.46	0.36	1961	2.13	1.74
1936	0.45	0.36	1962	2.39	1.94
1937	0.55	0.43	1963	2.64	2.12
1938	0.46	0.33	1964	2.99	2.40
1939	0.32	0.21	1965	3.36	2.68
1940	0.35	0.26	1966	4.31	3.46
1941	0.32	0.24	1967	3.96	3.11
1942	0.34	0.27	1968	4.65	3.70
1943	0.45	0.39	1969	6.28	5.04
1944	0.50	0.43	1970	6.19	4.95

l_D^* = expected losses on commercial bank deposits; an exponentially declining weighted average of current and past actual losses on deposits, using Friedman's consumption function weights ($\beta=.33$). When loss rates are not reported annually, but only as an average over five- or twenty-year periods annual estimates are obtained by interpolating by the annual number of bank suspensions (*FDIC Annual Reports, Banking and Monetary Statistics*).

s_{DD} = service charges on demand deposits (Cagan, *FDIC Annual Reports*).

All variables are annual averages centered on June 30. Friedman and Schwartz refer to data supplied by them.

REFERENCES

R. J. Barro and A. J. Santomero, "Household Money Holdings and The Demand Deposit Rate," *J. Money, Credit, Banking*, May 1962, *4*, 397–413.

M. Bronfenbrenner and T. Mayer, "Liquidity Functions in the American Economy," *Econometrica*, Oct. 1960, *28*, 810–34.

K. Brunner and A. H. Meltzer, "Predicting Velocity: Implications for Theory and Policy," *J. Finance*, May 1963, *18*, 319–54.

P. Cagan, *Determinants and Effects of Changes in the Stock of Money, 1875–1960*, New York 1965.

G. C. Chow, "On the Long-Run and Short-Run Demand for Money," *J. Polit Econ.*, Apr. 1966, *74*, 111–31.

C. F. Christ, "Interest Rates and Portfolio Selection Among Liquid Assets in the U.S.," in C. Christ et al., *Measurement in Economics*, Stanford 1963.

M. R. Darby, "The Allocation of Transitory Income Among Consumers' Assets," *Amer. Econ. Rev.*, Dec. 1972, *62*, 928–41.

R. Eisner, "Another Look at Liquidity Preference," *Econometrica*, July 1963, *31*, 531–38.

E. Feige, *The Demand for Liquid Assets: A Temporal Cross-Section Analysis*, New Jersey 1964.

———, "Expectations and Adjustments in the Monetary Sector," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 462–73.

M. Friedman, "The Quantity Theory of Money—A Restatement," in his *Studies in the Quantity Theory of Money*, Chicago 1956, 1–21.

———, "The Demand for Money—Some Theoretical and Empirical Results," in his *The Optimum Quantity of Money and Other Essays*, Chicago 1969, 41–134.

——— and **A. J. Schwartz**, *A Monetary History of the United States 1867–1960*, Princeton, 1963.

Y. Haitovsky, "A Note on the Maximization of $\bar{R}^2$," *Amer. Statist.*, Feb. 1969, *23*, 20–21.

H. R. Heller, "The Demand for Money—The Evidence from the Short-Run Data," *Quart. J. Econ.*, June 1963, *79*, 219–46.

B. Klein, "The Payment of Interest on Commercial Bank Deposits and the Price of Money: A Study of the Demand for Money," unpublished doctoral dissertation, Univ. Chicago, June 1970.

———, "The Demand for Quality-Adjusted Cash Balances," Nat. Bur. Econ. Res. unpublished paper, 1972.

———, "Income Velocity, Interest Rates and the Money Supply Multiplier: A Reinterpretation of the Long-Term Evidence," *J. Money, Credit, Banking*, May 1973, *5*, 656–68.

———, "The Recent Inflation and our New Monetary Standard," disc. pap. no. 48, UCLA, Mar. 1974.

D. Laidler (1966a) "Some Evidence on the Demand for Money," *J. Polit. Econ.*, Feb. 1966, *64*, 55–68.

———, (1966b) "The Rate of Interest and the Demand for Money—Some Empirical Evidence," *J. Polit. Econ.*, Dec. 1966, *64*, 545–55.

H. A. Latané, "Income Velocity and Interest Rates: A Pragmatic Approach," *Rev. Econ. Statist.*, Nov. 1960, *42*, 445–49.

T. H. Lee, "Alternative Interest Rates and the Demand for Money: The Empirical Evidence," *Amer. Econ. Rev.*, Dec. 1967, *57*, 1168–81.

———, "Alternative Interest Rates and the Demand for Money: Reply," *Amer. Econ. Rev.*, June 1969, *59*, 412–18.

A. H. Meltzer, "The Demand for Money: The Evidence from the Time Series," *J. Polit. Econ.*, June 1963, *17*, 219–46.

P. A. Samuelson, *Foundations of Economic Analysis*, Cambridge 1963.

R. T. Selden, "Monetary Velocity in the United States," in *Studies in the Quantity Theory of Money*, Chicago 1956, 179–257.

R. Teigen, "Demand and Supply Functions for Money in the United States: Some Structural Estimates," *Econometrica*, Oct. 1964, *32*, 476–509.

Board of Governors of the Federal Reserve System, *Fed. Res. Bull.*, various issues, 1961–71.

———, *Banking and Monetary Statistics*, Washington 1943.

———, "Member Bank Reserves and Related Items," *Supplement to Banking and Monetary Statistics, Sec. 10,* Washington 1962.

———, "Changes in Time and Savings Deposits and Commercial Banks," *Fed. Res. Bull.*, various issues, 1968–71.

Federal Deposit Insurance Corporation, (*FDIC*) *Annual Report*, Washington 1940, 1960–71.

———, *Assets and Liabilities of Commercial and Mutual Savings Banks*, June 1969-Dec. 1971.

The Effect of Risk on the Investment in Human Capital

By David Levhari and Yoram Weiss*

The theory of investment in human capital is conventionally developed under the assumption of perfect foresight. It is, of course, widely recognized that investment in human capital is subject to a considerable risk (see Gary Becker (1964, p. 55), Theodore Schultz (1971, p. 52), and Lester Thurow, pp. 71–74), but so far no attempt has been made to incorporate this element in the basic theory of human capital.

The purpose of this paper is to provide a *simple* theoretical framework in which the various effects of uncertainty may be analyzed. We consider a Fisherian two-period model where future labor earnings are randomly dependent on current investment in human capital. Similar models have been utilized by Hayne Leland, Agnar Sandmo, and Jacques Drèze and Franco Modigliani to analyze the problems of savings and portfolio choice under uncertainty. From the theoretical point of view, the special and perhaps novel feature of the present analysis is that the return of human capital depends on the amount invested. The average and marginal rates need not be equal and the correlation between them plays a central role in the analysis.

Within the simple two-period model, we examine the effect of uncertainty on the level of investment in human capital. In most of the analysis we focus on the case in which the returns from nonhuman capital are known with certainty. Implicit in this simplification is the notion that human capital is in some sense more risky than physical capital. From the social point of view, this need not be the case since compared with physical capital, human capital is quite flexible in its uses under varying economic circumstances. Nevertheless, from the point of view of each individual, human capital is probably more risky than physical capital. The main reason is that human capital cannot be bought or sold and cannot be separated from its owner. The possibilities for diversification are therefore very limited in the case of human capital. To diversify his human capital, the individual must acquire "general" education and forego the advantages of specialization. In the case of physical capital, the spread of ownership among many individuals allows specialization in production together with diversification of individual portfolios.

The hypothesis that human capital is more risky need *not* imply that the investment in human capital is discouraged or that the expected marginal return on human capital is higher. The relation between the expected marginal rate of return on human capital and on other assets depends upon the correlation between the marginal and average rates of return to human capital. If this correlation is positive, or, equivalently, if the variance in earnings is increasing with the level of schooling, then, and only then, will the expected return on human capital be higher than that of the safe nonhuman

* The Hebrew University of Jerusalem and Tel Aviv University, respectively. Financial support by the Foerder and the Falk Institutes is gratefully acknowledged. We wish to thank Peter Diamond, Theodore Schultz, Robert Jones, and the referee for their helpful comments.

asset. The hypothesis of increasing risk (i.e., increasing variance) is, of course, quite plausible and is shown to be consistent with empirical findings.

The simple two-period model is also used to analyze the comparative statics of investment in human capital under uncertainty. The effect of increased risk is shown to be ambiguous except for a special case. Under the assumptions of decreasing absolute risk aversion and increasing risk, it is shown that an increase in initial wealth will encourage the investment in human capital. We show that an increase in the rate of interest will induce a decrease in the investment in human capital when the individual is a net borrower during the investment period. When he is a net saver, an increase in the interest rate will lead to opposing income and substitution effects, and the result is ambiguous. We conclude with a rudimentary analysis of the effect of uncertainty under imperfect capital markets.

I. The Model

There are only two periods, the "present" and the "future." The individual planning involves decisions on the amounts invested in human and physical capital. We consider the investment in human capital to be time-consuming[1] (for example, schooling) and $0<\lambda<1$ is the proportion of the first period devoted to it. The proportion of the first period spent at work is $1-\lambda$. Utility depends only on present and future consumption; therefore investment in human capital in the last period (the future) is identically zero.

Future earnings y_1 depend on current investment in human capital λ and on the future (unknown) state of the world μ

$$y_1 = f(\lambda, \mu) \tag{1}$$

where μ is a random variable with a known distribution.

The random effect μ reflects two types of uncertainty which should perhaps be distinguished (see Marc Nerlove):

1) *Uncertain Inputs:* There is imperfect knowledge of individual exogenous characteristics such as ability which affect the earning capacity. There is also uncertainty of the quality of schooling; that is, the amount of human capital which would result from the commitment of time and money into the learning process.[2]

2) *Uncertain Output:* There is imperfect knowledge of future demand and supply conditions. Consequently the output (earning capacity) from a given level and type of human capital is uncertain.

The above distinction has bearing upon the specification of relation (1). For the uncertain input case, the assumptions $f_\mu(\lambda, \mu)>0$, $f_{\mu\mu}(\lambda, \mu)<0$, and perhaps also $f_{\lambda\mu}(\lambda, \mu)>0$ seem natural. On the other hand, if μ is merely an index of market conditions it becomes less clear which are the reasonable a priori restrictions. In this context even the seemingly harmless restriction $f_\mu(\lambda, \mu)>0$ is quite restrictive since it implies uniformity in the direction of supply and demand changes across all levels of skill.[3] If we assume that μ is just an ordinal index, then obviously there is not much meaning in restricting the second-order derivatives with respect to μ.

For each state of the world, the mar-

[1] In the more general case the output of human capital per unit of time is not fixed but depends on additional purchased inputs. However, the degree of substitution between these inputs is likely to be small. (See Becker (1967, pp. 5–9).)

[2] In fact, the length of the schooling period itself may be random since the ability to complete a given schooling program is also to some extent unknown. Similarly, the length of working life is uncertain. These types of uncertainty are not dealt with in the present paper.

[3] One could also consider *structural* random changes such as an increase in the demand for skilled workers together with a decrease in the demand for low skilled workers.

ginal productivity of schooling is assumed positive and decreasing ($f_\lambda(\lambda, \mu)>0$, $f_{\lambda\lambda}(\lambda, \mu)<0$).

The expected utility hypothesis is adopted. The individual's objective is then to maximize the expected value of lifetime utility from consumption.

$$\max_{c_0,\lambda} V = E\{u(c_0, c_1)\} \tag{2}$$

where $u(c_0, c_1)$, the utility function, is monotone and concave in both variables. Consumption in each period is decided upon after the income of that period is observed, and the individual obeys the wealth constraint:

$$c_1 = (A + (1 - \lambda)y_0 - c_0)(1 + r) + y_1 \tag{3}$$

where r is the market rate of interest, A is initial (inherited) wealth, and y_0 is the earning capacity in the first period. In general all these parameters may be random variables. However, in most of the discussion we assume A and y_0 to be known and exogenous to the individual.

Before we proceed to the description of the first-order conditions, it would be useful to suggest an alternative formulation which highlights the similarities and differences from the conventional dynamic portfolio model. We may rewrite equation (1) in the somewhat more special form

$$y_1 = y_0 + \lambda y_0 \gamma(\lambda, \mu) \tag{1'}$$

where λy_0 is the amount of investment in human capital in dollar terms, and $\gamma(\lambda, u)=(y_1-y_0)/\lambda y_0$ is the *average* rate of return for investment in human capital. Note that in (1′) uncertainty is endogenous (see Drèze and Modigliani) and can be eliminated if the level of investment in human capital is zero.

Equation (3) may be rewritten as

$$c_1 = K_0\left[\frac{\lambda y_0}{K_0}\gamma(\lambda, \mu) + \left(1 - \frac{\lambda y_0}{K_0}\right)(1 + r)\right] \tag{3'}$$

where $K_0=A+y_0+y_0/(1+r)-c_0$ is total investment out of the individual's "full income," and $A+y_0+y_0/(1+r)$ reflects wealth in the hypothetical case $\lambda=0$.

As seen in (3′), given the total portfolio K_0, λ determines its allocation between physical capital which yields $1+r$ and human capital which yields $\gamma(\lambda, \mu)$ per dollar invested. There are however two special features of the problem of investment in human capital which distinguish it from other portfolio models. The level of schooling only partially controls the structure of the individual's portfolio. Spending the same proportion of time at school implies different allocations of the portfolio depending upon its size. This is a reflection of the fact that human capital can be produced but cannot be bought or sold at the market. A second and related distinguishing feature is that the average return of human capital depends upon the level of investment.

The first-order conditions for the maximization of (2) subject to the constraints (1) and (3) are:

$$V_{c_0}=E\left\{\frac{\partial u}{\partial c_0}(c_0, c_1)-(1+r)\frac{\partial u}{\partial c_1}(c_0, c_1)\right\} = 0 \tag{4}$$

$$V_\lambda=E\left\{\frac{\partial u}{\partial c_1}(c_0, c_1)(-y_0(1+r) + f_\lambda(\lambda, \mu))\right\} = 0 \tag{5}$$

Condition (4) states that no gain can be achieved by transferring consumption from period to period. When the market interest rate r is known with certainty, (4) reduces to the requirement that the ratio between the expected future and present marginal utilities equal the market discount factor.

We assume that $\partial u(0, c_1)/\partial c_0$ and $\partial u(c_0, 0)/\partial c_1$ are both "large" so that the individual always wishes to consume at a positive level. In an uncertain world this implies

that current consumption will be less than $A+(1-\lambda)y_0+f(\lambda,\underline{\mu})/(1+\bar{r})$ where $\underline{\mu}$ is the known lower bound on μ and $\bar{r}$ is the known upper bound on the market interest rate. The individual will borrow at most against the *minimal* present value of his future labor earnings. Condition (5) states that at the optimum the expected *marginal* (in utility terms) rates of return from physical and human capital must be equal. For simplicity, we assume that an interior solution such that $0<\lambda<1$ is attained.

Conditions (4) and (5) together with (1) and (3) determine the optimal values of present consumption and investment in human capital which we denote by c_0^* and λ^*, respectively. Future consumption c_1 is, of course, a *random* variable whose distribution depends on the distribution of y_1 and on the choice of c_0 and λ. Our assumptions on the utility and production functions ensure that V is concave in λ and c_0.[4] Conditions (4) and (5) are therefore sufficient for maximum and their solution must be unique.

II. The Expected Marginal Rate of Return from Human Capital

Our first purpose is to compare the expected rate of return from human capital with that of physical capital at the individual optimum. As we shall see, when the expected marginal rate of return from human capital exceeds that of physical capital, it is indicated that the individual views human capital as more risky on the *margin*. That is, a marginal increase in the level of investment in human capital increases the variance of future earnings and consumption.

If one assumes that μ is an independent random variable across individuals,[5] and if subjective and objective probabilities coincide, then for a discussion of over- or underinvestment in human capital it is the difference between the expected marginal rates of return which is relevant. If the expected rate of return on human capital is above $1+r$ (or $E(1+r)$), then it is desirable from a social point of view to transfer resources from physical into human capital investment. Since the investment is performed by many individuals, society will enjoy the average (expected) rate of return.

Condition (5) may be rewritten as

$$(6) \quad E\left\{\frac{\partial u}{\partial c_1}(c_0^*, c_1)\right\}[E\{f_\lambda(\lambda^*,\mu)\} - E\{(1+r)\}y_0] = \text{cov}\left\{\frac{\partial u}{\partial c_1}(c_0^*, c_1), r\right\}y_0 - \text{cov}\left\{\frac{\partial u}{\partial c_1}(c_0^*, c_1), f_\lambda(\lambda^*,\mu)\right\}$$

Suppose, initially, that the market interest rate is known with certainty, it then follows from (6) that[6]

$$(7) \qquad E\left\{\frac{f_\lambda(\lambda^*,\mu)}{y_0}\right\} \gtreqless 1+r$$

if, and only if, $f_{\lambda\mu}\gtreqless 0$. (We assume $f_\mu>0$, and that, for a given λ, $f_{\lambda\mu}$ is of the same sign for all μ.)[7]

The difference between the expected marginal rates of return on human and physical capital is thus seen to depend

[4] Let $\psi(c_0, \lambda, \mu)=u[c_0, (A+(1-\lambda)y_0-c_0)(1+r)+f(\lambda, \mu)]$. Then $\psi_{c_0c_0}=u_{00}-2u_{01}(1+r)+(1+r)^2u_{11}<0$, $\psi_{\lambda\lambda}=u_{11}[-y_0(1+r)+f_\lambda]^2+u_1f_{\lambda\lambda}<0$ and $\psi_{\lambda c_0}=[-y_0(1+r)+f_\lambda][u_{10}-(1+r)u_{11}]$. It is easily seen that $\psi_{c_0c_0}\psi_{\lambda\lambda}>[\psi_{c_0\lambda}]^2$. Since $\psi(c_0, \lambda, \mu)$ is concave in c_0 and λ for all μ, $V=E\{\psi(c_0, \lambda, \mu)\}$ must also be concave.

[5] The assumption that μ is independent across individuals is rather restrictive. It may be valid in the case of uncertain inputs (say, ability), and perhaps also in the case of output uncertainty provided that different occupations are affected independently. See Nerlove.

[6] Given the assumptions that $f_\mu>0$ and $\partial^2u/\partial c_1^2<0$, $f_{\lambda\mu}\gtreqless 0$ implies that $\text{cov}\{\partial u/\partial c_1, f_\lambda\}\lesseqgtr 0$.

[7] A special case of some interest is when $f_{\lambda\mu}=0$ uniformly or equivalently $f(\lambda, \mu)=\phi(\lambda)+\mu$. It is seen immediately from the first-order condition (5) that $\phi'(\lambda)/y_0=1+r$ for *all* states of the world. Uncertainty in this additive formulation has no effect on the production decisions of the individual. Consumption decisions, on the other hand, will be affected.

upon whether the marginal product of human capital is increasing or decreasing with μ.

It is worth noting that in contrast to standard portfolio analysis, risk aversion by itself is *not* sufficient to ensure that for an interior solution $E\{f_\lambda(\lambda^*, \mu)/y_0\} > 1+r$; that is, that the expected marginal rate of return from the risky asset is higher. It is necessary to add the requirement of increasing risk ($f_{\lambda\mu}>0$). The source of the difference is the dependence in our model of the marginal rate of return on the level of investment. This dependence allows the possibility of a reduction in the variance of future earnings as the level of schooling increases.

Comparisons may also be made between the expected *average* rate of return from human capital and the market interest rate. It can be shown[8] that if all uncertainty is endogenous as in (1′), then at the optimum, the expected average rate of return from human capital will exceed that of nonhuman capital. In this case the assumption that $f_{\lambda\mu}>0$ uniformly is not needed.

In the more general case in which r is also random, only weaker results are obtained. It can be shown,[9] for instance, that if the correlation between μ and r is nonpositive, and if $f_{\lambda\mu}$ is positive, and finally if the individual is a net borrower, then the expected marginal rate of return from human capital exceeds the expected rate of return from physical capital.

The economic meaning of proposition (7) and the special role of $f_{\lambda\mu}$ is more easily grasped with the aid of Figure 1. Suppose there are only two states μ_1 and μ_2. Let μ_2 represent a situation of relatively high demand. In both Figures 1a and 1b the *average* productivity of human capital is assumed to be higher at state μ_2 for all levels of λ. (The convention $f_\mu>0$ implies that the assigned indexes satisfy $\mu_2>\mu_1$.) The *marginal* productivity of investment in human capital may nevertheless decrease or increase. It will decrease if $f_{\lambda\mu}<0$ and increase if $f_{\lambda\mu}>0$. Thus a positive (negative) sign of $f_{\lambda\mu}$ corresponds to a positive (negative) correlation between the *marginal* and the *average* rates of return.

Note that if uncertainty is endogenous, that is, $f(0, \mu)=y_0$, as in (1′), then $f_{\lambda\mu}$ must be positive at least around $\lambda=0$. For $f_{\lambda\mu}$ to be negative uniformly it is necessary that there also be uncertainty of future earnings for low skilled workers.

It is important to note that the sign of $f_{\lambda\mu}$ can be determined empirically. A direct but partial test would concentrate on the complementary relationship between measured ability and schooling. Findings by John Hause indicate that a given increase in ability adds more to the earnings of the relatively educated. A more general but indirect test may be derived from the observation that a positive correlation between the marginal and average rates of return (i.e., between f_λ and y_1) exists, if and only if, the variance of earnings increases with the level of investment in human capital.[10] Indeed there are some empirical findings which indicate that the variance (and of course the mean) of the

[8] Due to the concavity of $u(c_0, c_1)$, $u(c_0^*, (A+(1-\lambda^*)y_0-c_0^*)(1+r)+E\{f(\lambda^*, \mu)\}) > E\{u(c_0^*, (A+(1-\lambda^*)y_0-c_0^*)(1+r)+f(\lambda^*, \mu))\} > E\{u(c_0^*, (A+y_0-c_0^*)(1+r)+f(0, \mu))\}$ since by definition c_0^*, λ_0^* maximize expected utility. Suppose that $f(0, \mu)=y_0$ and is thus nonrandom, $u(c_0^*, (A+(1-\lambda^*)y_0-c_0^*)(1+r)+E\{f(\lambda^*, \mu)\}) > u(c_0^*, (A+y_0-c_0^*)(1+r)+y_0)$. It follows that $E\{f(\lambda^*, \mu)\}-y_0/\lambda^* y_0 > 1+r$.

[9] Examining the signs of the covariances which appear in (6), we note that: *sign* cov $\{\partial u(c_0^*, c_1)/\partial c_1, r\} = -$*sign* cov $\{c_1, r\}$ and cov $\{c_1, r\} =$ cov $\{s^*(1+r)+f(\lambda^*, \mu), 1+r\}$, where $s^*=A+(1-\lambda^*)y_0-c_0^*$. Thus cov $\{c_1, r\} = s^*\sigma^2_{(1+r)} +$ cov $\{f(\lambda^*, \mu), 1+r\}$. Similarly *sign* cov $\{\partial u(c^*, c_1)/\partial c_1, f_\lambda(\lambda^*, \mu)\} = -$*sign* cov $\{c_1, f_\lambda(\lambda^*, \mu)\}$ and cov $\{c_1, f_\lambda\} = s^*$ cov $\{1+r, f_\lambda\} +$ cov $\{f_\lambda, f\}$. For $f_{\lambda\mu}>0$, cov $\{f_\lambda, f\}>0$. Assuming cov $\{\mu, r\} \le 0$ and $s^* \le 0$, we have cov $\{\partial u/\partial c_1, r\} \ge 0$ and cov $\{\partial u/\partial c_1, f_\lambda\} \le 0$ so that $E\{f_\lambda\{\lambda^*, \mu\} - E\{(1+r)y_0\} \ge 0$.

[10] $\partial/\partial\lambda$ var $\{f(\lambda, \mu)\} = E\{\partial/\partial\lambda[f(\lambda, \mu)-E\{f(\lambda, \mu)\}]^2\} = 2$ cov $\{f(\lambda, \mu), f_\lambda(\lambda, \mu)\}$. The reader is also referred to Figures 1a and 1b, where it is seen that $f_{\lambda\mu}>0$ implies that $\sigma^2_{y_1}$ increases with λ, and vice versa.

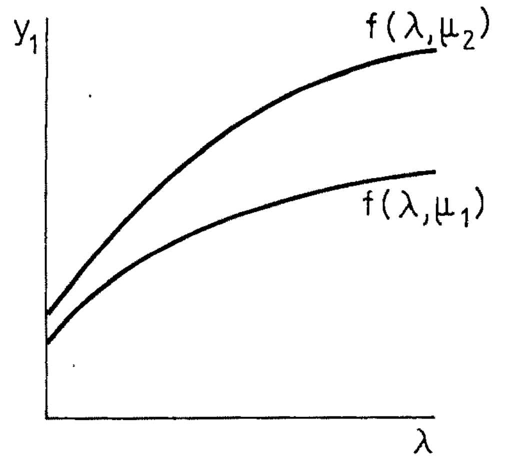


FIGURE 1a: $f_{\lambda\mu}>0$

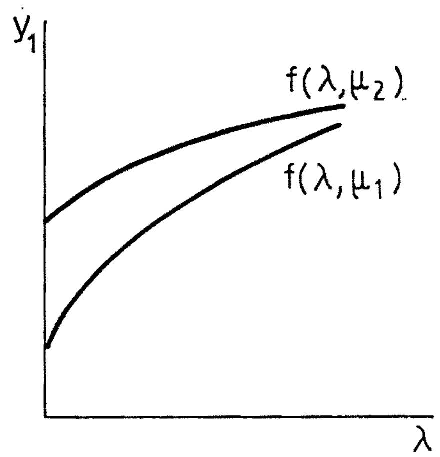


FIGURE 1b: $f_{\lambda\mu}<0$

earning distribution[11] tends to increase with the level of schooling. It is perhaps worth mentioning that such a tendency is not necessarily obvious or "natural." Burton Weisbrod, for instance, argues that $f_{\lambda\mu}$ is likely to be negative since more educated individuals can more easily switch from occupations which become unexpectedly obsolete.

III. The Optimal Level of Investment

The results of the previous section allow some comparisons between the level of investment in human capital under conditions of uncertainty with the level of investment which would be optimal in the hypothetical case of certainty. From the point of view of the individual "certainty" may have two different meanings. 1) The individual may be insured in such a way that his earnings for each level of schooling are independent of the state of the world,[12] or 2) the individual may have prior information on the state of the world which will be obtained.

To avoid meaningless comparisons, some critical and relevant value must be the *same* in both situations. We shall consider the following comparisons:

1) $y_1=E(y_1)$ or $g(\lambda)=E\{f(\lambda,\ \mu)\}$ where $g(\lambda)$ is the future earnings function of the *insured* individual.

2) The state of the world is known, $\mu=\mu_0$ where μ_0 is chosen so as to equate the marginal productivity of investment in human capital under certainty with the *expected* marginal productivity under uncertainty.

$$E\{f_\lambda(\lambda^*,\ \mu)\} = f_\lambda(\lambda^*,\ \mu_0) \tag{8}$$

Alternatively we could hold the expected *average* return the same in both situations. Due to the special features of our problem, namely the dependence of

[11] See Hendrick Houthhakker, Table 1 and Weiss (1972), Table 1. These studies do not report the variances directly but they can be calculated from the means and the coefficients of variation which are reported. In both studies the variances in almost all age groups tend to increase with the level of schooling. Weiss (1972) also reports a tendency of the coefficient of variation to *decrease* with the level of schooling. This suggests an added restriction on $f(\lambda,\ \mu)$ namely cov $\{f(\lambda,\mu), f_\lambda(\lambda,\mu)\}E\{f(\lambda,\mu)\} < \text{var}\ \{f(\lambda,\mu)\}E\{f_\lambda(\lambda,\mu)\}$. A simple form of $f(\lambda,\ \mu)$ which may be useful in actual applications suggests itself: $f(\lambda,\ \mu)=\mu\phi(\lambda)$. This multiplicative form may be viewed as a borderline case where the variance is increasing with λ while the coefficient of variation remains unchanged.

[12] It is somewhat unlikely that a market for such an insurance will be established, due to the moral hazards involved. Notice, however, that some of the moral hazards are avoided since the insurance plan is such that for each level of investment in human capital the individual is assured of expected value income associated with that level.

the rate of return on the level of investment, such a definition would provide a different value to μ_0. Needless to say it is the *marginal* value which is relevant to the choice of the optimal level of education.

3) $\mu_0 = E\{\mu\}$. Under this definition the average value of the random variable μ is held at the same level in both situations. This definition is somewhat questionable in the case in which μ is merely an ordinal index of market conditions.

Note that μ_0 will be the same under the two last definitions if μ appears in the production function (1) in a *multiplicative* form (that is, $f(\lambda, \mu) = a + \mu\phi(\lambda)$).

1) Given $y_1 = g(\lambda)$ the individual will invest under certainty up to the point where:

$$\text{(9)} \qquad \frac{g'(\lambda_1^0)}{y_0} = 1 + r$$

By definition $g'(\lambda_1^0) = E\{f_\lambda(\lambda^0, \mu)\}$. It follows from (7) and $f_{\lambda\lambda} < 0$ that

$$\text{(10)} \qquad \lambda^* \lesseqgtr \lambda_1^0 \quad \text{if, and only if,} \quad f_{\lambda\mu} \gtreqless 0$$

That is, a positive (negative) correlation between the average and marginal returns for human capital implies that the insured individual will invest less (more) in human capital under conditions of uncertainty.

2) Given μ_0 as defined by (8) the individual will invest under certainty up to the point at which

$$\text{(9')} \qquad \frac{f_\lambda(\lambda_2^0, \mu_0)}{y_0} = 1 + r$$

By definition

$$\frac{f_\lambda(\lambda^*, \mu_0)}{y_0} = E\frac{\{f_\lambda(\lambda^*, \mu)\}}{y_0}$$

Again it follows from (7) and from the assumption of decreasing marginal productivity that

$$\text{(10')} \qquad \lambda^* \lesseqgtr \lambda_2^0 \quad \text{if, and only if,} \quad f_{\lambda\mu} \gtreqless 0$$

3) Given $\mu_0 = E(\mu)$ the individual will invest under certainty up to the point at which

$$\text{(11)} \qquad \frac{f_\lambda(\lambda_3^0, E\{\mu\})}{y_0} = 1 + r$$

If the marginal productivity of human capital is *concave* in μ then it follows from (7) that increasing risk ($f_{\lambda\mu} > 0$) implies that $\lambda^* < \lambda_3^0$. On the other hand, if $f_\lambda(\lambda, \mu)$ is strictly convex in μ, $\lambda^* > \lambda_3^0$ need not follow.

The intuitive explanation of the above results and of the special role played by $f_{\lambda\mu}$ is quite clear. The contribution of an additional unit of investment in human capital is weighted by the marginal utility of consumption at each state of the world. Positive correlation between f_λ and y_1 implies that states with higher marginal rates of return from human capital are given relatively lower weight. Or, put differently, an increase in λ implies in this case that the variance of future earnings is increased. The investment in human capital is therefore discouraged. Negative correlation between f_λ and y_1 has the opposite effect. An increase in the level of schooling will reduce the variance of future earnings. The individual will hedge against states with low earnings by increasing his investment in human capital.

Our analytical discussion has brought us to an uncertain conclusion. Somewhat surprisingly, uncertainty may encourage investment in human capital and, of course, uncertainty may discourage investment. If we accept on empirical grounds the hypothesis of increasing risk, that is, $f_{\lambda\mu} > 0$, then it follows that a maximizing individual will reduce his investment in human capital under conditions of uncertainty.

IV. Comparative Statics

We now turn to the comparative statics of investment in human capital under un-

certainty. In particular we examine the effects of: a) increase in initial wealth; b) increase in the market rate of interest; and c) increased risk.

Throughout this discussion we shall assume that the market interest rate is known with certainty. For simplicity we shall also assume that the utility function is separable $(u(c_0,c_1)=u_0(c_0)+u_1(c_1))$ so that present and future consumptions are independent goods.

Under our assumption of decreasing marginal utility and decreasing marginal productivity of human capital, $V(c_0, \lambda)$ is concave. We thus have $V_{\lambda\lambda}<0$, $V_{c_0c_0}<0$ and $V_{cc}V_{\lambda\lambda}>[V_{c\lambda}]^2$. Furthermore it can be shown[13] that at the optimum $V_{c_0\lambda}<0$ provided that:

1) $f_{\lambda\mu}>0$ (principle of increasing risk).

2) $u_1''(c_1)/u_1'(c_1)$ is increasing (decreasing absolute risk aversion). In the following analysis we assume these two conditions to be satisfied.

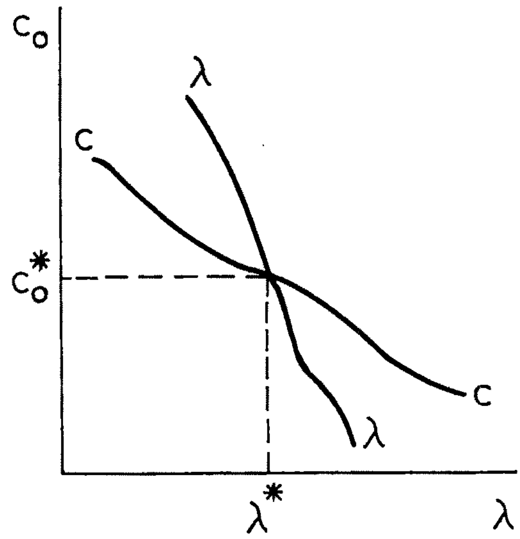


FIGURE 2

In the analysis of comparative statics it will be helpful to utilize Figure 2. The curves cc and $\lambda\lambda$ in Figure 2 indicate the combinations of c_0 and λ which satisfy conditions (4) $V_{c_0}=0$, and (5) $V_\lambda=0$, respectively. From the remarks in the previous paragraph, it follows that at the point of intersection both curves are negatively sloped and the slope of $\lambda\lambda$ is larger in absolute value.

Consider first the effect of an increase in initial wealth A. A somewhat surprising aspect of the traditional theory of human capital is that under perfect capital markets, the amount invested and, in particular, the level of schooling is independent of initial wealth. This has led to attacks on the human capital approach for ignoring the consumption aspects of schooling (see Harry Shaffer). This criticism is met, at least partially, in the present model. The introduction of risk breaks the separation between consumption and production decisions, and the level of investment is therefore related to initial wealth.

An increase in A will shift the $\lambda\lambda$ curve in Figure 2 upwards. An increase in c_0 by the same magnitude as in A will restore equality in (5). The cc schedule will also shift upwards but by a lesser degree. (Due to decreasing marginal utility an equal increase in c_0 and A will lead to a negative value of V_{c_0}.) The result is therefore that

[13] $V_{c_0\lambda}=-(1+r)E\{u_1''(c_1)[-y_0(1+r)+f_\lambda(\lambda^*, \mu)]\}=(1+r)E\{R_A(c_1)u_1'(c_1)[-y_0(1+r)+f_\lambda(\lambda^*, \mu)]\}$ where R_A denotes the absolute degree of risk aversion. Let $g(\mu)$ be the density function of μ and denote μ' the state at which $-y_0(1+r)+f_\lambda(\lambda^*, \mu)=0$. Then

$$E\{R_A u_1'(c_1)[-y_0(1+r)+f_\lambda]\}$$
$$=\int_{\underline{\mu}}^{\mu'} R_A(c_1)u_1'(c_1)[-y_0(1+r)+f_\lambda]g(\mu)d\mu$$
$$+\int_{\mu'}^{\bar{\mu}} R_A(c_1)u_1''(c_1)[-y_0(1+r)+f_\lambda]g(\mu)d\mu$$

Since $f_{\lambda\mu}>0$ the first integral is negative and the second is positive. Since $R_A(\ )$ is monotonically decreasing in c_1 and thus in μ, we get $E\{R_A u_1'(c_1)[-y_0(1+r)+f_\lambda]\}<R_A(c_1')E\{u'(c_1)[-y_0(1+r)+f_\lambda]\}$, where c_1' denotes the level of future consumption at state μ'. From the first-order condition (5) the right-hand side of this inequality is zero. Hence $V_{c_0\lambda}<0$.

the equilibrium level of investment in human capital must increase.[14]

In light of our assumptions, this result should have been expected. Decreasing absolute risk aversion implies that risky income is a normal good. The wealthier individual can afford more risk. Since risk is increasing with the level of investment in human capital, investment is encouraged. It can further be shown[15] that for a fixed c_0 any increase in absolute risk aversion, not necessarily that which results from increasing wealth, will reduce the investment in human capital. However, if c_0 is allowed to vary λ^* need not be reduced.

Consider now the effect of a change in the interest rate. Under conditions of certainty an increase in r will unambiguously discourage the investment in human capital. This result is obtained in an uncertain world only if the individual is a net borrower in the first period. When the individual is a net saver, the effect of an increase in r is ambiguous.

The effect of an increase in the interest rate on investment in human capital can be broken into two parts:[16] a substitution effect of a determined (negative) sign and an income effect which depends on whether the individual is a net borrower or net saver. The negative sign of the substitution effect reflects the tendency to reduce the investment in human capital when the opportunity costs (the foregone returns from savings) increase. The appearance of an income effect reflects again the fact that, under uncertainty, changes in wealth affect the production decisions of the individual. If the individual is a net borrower during the first period $(A+(1-\lambda^*)y_0 - c_0^*<0)$, an increase in the interest rate reduces his present consumption opportunities. Since investment in human capital increases with wealth, this decrease in wealth will discourage investment in human capital. The income and substitution effects being of equal sign implies that, in the case of a net borrower, an increase in interest rate will unambiguously lead to a reduction in the investment in human capital. On the other hand, when the individual is a net saver $(A+1(1-\lambda^*)y_0 - c_0^*>0)$, the income and the substitution effects have opposite signs and the net

[14] $$|\Delta| \frac{d\lambda^*}{dA} = V_{c_0c_0}(-V_{\lambda A}) - V_{c_0\lambda}(-V_{c_0A})$$

where
$$V_{c_0c_0} = u_0''(c_1) + (1+r)^2E\{u_1''(c_1)\}$$
$$V_{c_0A} = -(1+r)^2E\{u_1''(c_1)\}$$
$$V_{\lambda A} = (1+r)E\{u_1''(c_1)[-y_0(1+r)+f_\lambda]\} = -V_{c_0\lambda}$$

and $$|\Delta| = V_{c_0c_0}V_{\lambda\lambda} - [V_{c_0\lambda}]^2 > 0$$

Hence
$$|\Delta| \frac{d\lambda^*}{dA} = V_{c_0\lambda}(V_{c_0c_0} + V_{c_0A}) = V_{c_0\lambda}u_0\ (c) > 0$$

[15] Denote μ_1 and c_1' as in fn. 13. Dividing the first-order condition (5) by $u_1'(c_1')$ we can rewrite it as

$$V_\lambda = \int_{\underline{\mu}}^{\mu'} \frac{u_1'(c_1)}{u_1'(c_1')}[-y_0(1+r)+f_\lambda(\lambda^*,\mu)]g(\mu)d\mu + \int_{\mu'}^{\bar{\mu}} \frac{u_1'(c_1)}{u_1'(c_1')}[-y_0(1+r)+f_\lambda(\lambda^*,\mu)]g(\mu)d\mu = 0$$

Greater risk aversion increases $u'(c_1)/u'(c_1')$ for $\mu\in[\underline{\mu},\mu']$ and decreases it for $\mu\in[\mu',\bar{\mu}]$ (see John Pratt, pp. 128–29). Thus, higher risk aversion implies that, at λ^*, V_λ is negative. A reduction in the investment in human capital is necessary to restore the equality in (5).

[16] $$|\Delta| \frac{d\lambda^*}{dr} = V_{c_0c_0}(-V_{\lambda r}) - V_{c_0\lambda}(-V_{c_0r})$$

where
$$V_{\lambda r} = E\{-u_1'(c_1)y_0 + u_1''(c_1)S^*[-y_0(1+r)+f_\lambda]\}$$
$$V_{c_0r} = -(1+r)E\{u_1''(c_1)\}S^*$$

and $$S^* = A + (1-\lambda^*)y_0 - c_0^*$$

Note that $C_{c_0r} = (S^*/1+r)V_{c_0A}$ (see fn. 14) and $V_{\lambda r} = -y_0E\{u_1'(c_1)\} + (S^*/1+r)V_{\lambda A}$ hence $d\lambda^*/dr = (S^*/1+r)(d\lambda^*/dA) + y_0V_{c_0c_0}E\{u_1'(c_1)\}|\Delta|^{-1}$. The first term on the right-hand side is the income effect weighted by the amount of saving at the first period. The second term, the substitution effect, is always negative.

effect of an increase in r on λ^* is therefore ambiguous.

Consider, finally, the problem of increase in risk. If one adopts the definition of increased risk suggested by Michael Rothschild and Joseph Stiglitz, then the marginal effect of risk depends on whether ψ_{c_0} and ψ_λ are concave or convex in μ. (We define $\psi(c_0, \lambda, \mu)=u_0(c_0)+u_1((A+(1-\lambda)y_0-c_0)(1+r)+f(\lambda, \mu))$ so that $V_\lambda=E\{\psi_\lambda\}$ and $V_{c_0}=E\{\psi_{c_0}\}$.) An increase in risk will be reflected in a downward (upward) shift of the cc and $\lambda\lambda$ curves in Figure 2 when ψ_{c_0} and ψ_λ are concave (convex). There are two difficulties in applying this procedure to the present model:

1) Since c_0 and λ are substitutes, $V_{c_0\lambda}<0$. Qualitative information on the direction of shifts is not sufficient to determine the final outcome; the size of the shifts matters.

2) The concavity or convexity of ψ_λ *cannot* be determined on an a priori basis (except in the special case in which the utility function is quadratic and μ has a multiplicative effect).[17]

We shall therefore adopt a narrower definition of increased risk. Following Kenneth Arrow and Sandmo, we shall consider multiplicative shifts in the distribution of μ holding its mean constant. To simplify further, assume that the random effect which appears in (1) has a multiplicative effect. That is, $f(\lambda, \mu)=a+\mu\phi(\lambda)$ with $\phi'>0$, $\phi''<0$, and $E\{\mu\}=1$. We have already shown that the mere existence of uncertainty, that is, an increase in risk from zero to a positive level, will in this case reduce investment in human capital. (Note that the hypothesis of increasing risk, i.e., $f_{\lambda\mu}>0$, is satisfied in the multiplicative case.) However, it is also possible to show (see the Appendix) that in case of a multiplicative random effect, *any* multiplicative shift in the distribution of μ will reduce the investment in human capital.

V. Some Special Cases

Our purpose in this section is to examine the effects of some changes in the assumptions of the basic model. Specifically we examine the possibilities that a) earning power in the first period, y_0, is random, and b) the capital market is imperfect.

It was assumed in the previous sections that the alternative costs during the first period are known with certainty. It is quite possible, however, that at the points at which the decisions to enter school or to leave it are made, earnings opportunities out of school are uncertain (see Becker (1964, p. 108)). Consumption decisions are also typically made before y_0 is realized.

Assuming that the market rate of interest is known with certainty, the first-order condition (5) can be written as follows:

$$(6')\quad E\left\{\frac{\partial u(\overset{*}{c_0}, c_1)}{\partial c_1}\right\} E\{-y_0(1+r)+f_\lambda\} = (1+r)\operatorname{cov}\left\{\frac{\partial u(\overset{*}{c_0}, c_1)}{\partial c_1}, y_0\right\} - \operatorname{cov}\left\{\frac{\partial \overset{*}{u}(c_0, c_1)}{\partial c_1}, f_\lambda\right\}$$

[17]
$$\frac{\partial^2\psi_{c_0}}{\partial\mu^2} = -[u_1'''(c_1)f_\mu^2+u_1''(c_1)f_{\mu\mu}](1+r)$$
$$\frac{\partial^2\psi_\lambda}{\partial\mu^2} = [-y_0(1+r)+f_\lambda][u_1'''(c_1)f_\mu^2+u_1''(c_1)f_{\mu\mu}] + 2u_1''(c_1)f_{\lambda\mu}f_\mu+u_1'(c_1)f_{\lambda\mu\mu}$$
Assuming decreasing absolute risk aversion and $f_{\mu\mu}\le 0$ implies $\partial^2\psi_{c_0}/\partial\mu^2<0$. Thus for given λ increase in risk will reduce present consumption. Unfortunately very little can be said on the sign of $\partial^2\psi_\lambda/\partial\mu^2$, and thus on the general effect on λ^* and c^*. Under the quadratic utility function $u'''=0$. Under the multiplicative assumption $f_{\lambda\mu}>0$ and $f_{\mu\mu}=f_{\lambda\mu\mu}=0$. Hence $\partial^2\psi_{c_0}/\partial\mu^2=0$ and $\partial^2\psi_\lambda/\partial\mu^2<0$. Referring to Figure 2, we note that under the quadratic utility assumption, absolute risk aversion is *increasing* at the relevant range. The curves cc and $\lambda\lambda$ are therefore upward sloping. An increase in risk will not affect cc but $\lambda\lambda$ will shift upwards and to the left. Both c_0 and λ will decrease.

Both covariances on the right-hand side are likely to be negative in the realistic case where present and future earnings are positively correlated.[18] It is thus in general impossible to determine whether the expected return on investing extra time in human capital $E\{f_\lambda(\lambda, \mu)\}$ is above its expected alternative cost $(1+r)E\{y_0\}$. To take an extreme example, suppose that $y_0 = \mu\bar{y}_0$ and $y_1 = \mu\phi(\lambda)$ so that y_0 and y_1 are perfectly correlated. It is immediately seen from the first-order condition (5) that in this case uncertainty has no effect on investment in human capital since λ^* is the same for *all* states of the world. This may be the case if unknown ability (to earn money) is the only source of uncertainty (see Weiss (1971a)).

We conclude with some brief comments on the case of imperfect capital markets. An extreme but fairly reasonable (in the face of uncertainty of future earnings) form of imperfection is such that borrowing against future earnings is not feasible. The individual is now operating under the additional constraint

$$c_0 \leq A + (1 - \lambda)y_0 \tag{12}$$

Future consumption, as before, is determined by (3) where r is now the *lending* rate. Two situations may occur:

1) The no-borrowing constraint (12) is ineffective in which case the previous analysis still holds;

2) Constraint (12) is binding so that c_0 and λ become linearly dependent on each other. Suppose now that (12) is in fact binding. Instead of a joint portfolio-saving problem, the problem now becomes simply that of saving in one asset —human capital.

The first-order condition assumes the form

$$-u_0'(c_0)y_0 + E\{u_1'(c_1)f_\lambda(\lambda, \mu)\} = 0 \tag{13}$$

(For simplicity we retain the assumption of a separable utility function.)

The interpretation of (13) is quite clear. An increase in the investment in human capital implies less earnings and thus less current consumption. At the optimum this marginal cost (in utility terms) must be equal to the expected future contribution of a unit of schooling. In the special case in which utility is of the form $u_0(c_0)+u_1(c_1) = u(c_0)+(1+\rho)^{-1}u(c_1)$, where ρ is a fixed subjective discount rate for future utilities, condition (13) reduces to the requirement that the expected rate of return in utility terms must equal the subjective discount factor.[19]

It is easily shown that in this case an increase in initial wealth A will increase the optimal investment in schooling.[20] In contrast to the case of a perfect capital market, risk aversion as such is sufficient to ensure this result. Of course, the same result will also hold under conditions of certainty when debt limitations are binding.

The fact that family wealth has a positive effect on the investment in schooling is widely recognized (see Samual Bowles), but typically this phenomenon is ascribed to either an imperfect capital market (see Becker (1967, pp. 12–13)) or to the consumption benefits associated with schooling (see Thurow, p. 80 and Shaffer). As we have seen in this paper, uncertainty of future earning is an additional cause for such a relation. Of course, from the point of view of testable hypotheses it is difficult

[18] Notice that *sign* cov $\{\partial u(c_0^*, c_1)/\partial c_1, y_0\} = -$*sign* cov $\{c_1, y_0\}$ and *sign* cov $\{\partial u(c_0^*, c_1)/\partial c_1, f_\lambda(\lambda^*, \mu)\} = -$*sign* cov $\{c_1, f_\lambda(\lambda^*, \mu)\}$. Now, cov $\{c_1, y_0\} =$ cov $\{(A+y_0(1-\lambda)-c_0)(1+r)+f(\lambda^*, \mu), y_0\} = (1-\lambda^*)(1+r)\sigma^2 y_0 +$ cov $(f(\lambda^*, \mu), y_0) > 0$ provided that cov $\{\mu, y_0\} > 0$. Also cov $\{c_1, f_\lambda\} = (1-\lambda^*)(1+r)$ cov $\{y_0, f_\lambda\} +$ cov $\{f(\lambda^*, \mu), f_\lambda(\lambda^*, \mu)\} > 0$ since $f_{\lambda\mu}$ is assumed positive.

[19] For an attempt to estimate such rates of return empirically, see Weiss (1972).

[20] $$\frac{d\lambda^*}{dA} = -\frac{-u_0''(c_0)y_0}{u_0''(c_0)y_0^2 + E\{u_1''(c_1)f_\lambda^2 + u_1'(c_1)f_{\lambda\lambda}\}} > 0$$

to discriminate between the situations of uncertainty and imperfect capital markets.

One may also examine the effect of increased risk under uncertainty and (strictly) imperfect capital markets. As is well known (see Rothschild and Stiglitz), the result of such a change on the level of investment in a single asset is in general ambiguous. When the random effect in the production function assumes a multiplicative form, the effect of increased risk depends only on the shape of the utility function and in particular on the relative degree of risk aversion.[21]

VI. Conclusions

The various examples which we considered in this paper show that "uncertainty matters." The introduction of uncertainty has a significant effect on the testable hypotheses which one would derive from theory of human capital, as well as on the interpretation and policy implications of existing evidence such as estimates of rates of return. Not surprisingly, the nature of the modifications which must be made when risk is introduced cannot be determined on an a priori level. It is necessary to presuppose some statistical relations which need to be verified empirically. Specifically we noticed the importance of the correlations between the average and marginal returns of schooling, between the returns of human and nonhuman capital, and between current and future earnings (see Becker (1964, pp. 104–13)).

The present analysis may be extended in various ways, two of which are of some interest:

1) A joint occupational-educational choice model with different *types* of human capital.

2) A multiperiod model in which the investment in human capital is distributed over time.[22]

Within the generalized model important questions arise which we did not touch. For instance, under certainty it is profitable to specialize in one occupation and to concentrate the investment in schooling at an early age (see Ben-Porath and Weiss (1971b)). Will these tendencies hold also when future earnings are uncertain? Another issue is the time pattern of the random future income streams. One would suspect, for instance, that the individual will prefer to receive a given risk at an early age.[23]

Appendix

In this Appendix we show that a multiplicative shift decreases the investment in human capital provided that the production function is such that f_μ depends on λ alone; that is, $f(\lambda, \mu) = a + \mu\phi(\lambda)$. Without loss of generality, we may assume $E\{\mu\} = 1$. Define $\mu(h) = \mu + h(\mu - 1)$. From the first-order conditions:

[21] The second-order derivative of $u_1'(c_1)f_\lambda(\lambda, \mu)$ with respect to μ is given by $u_1'''f_\mu^2 f_\lambda + u_1''f_{\mu\mu}f_\lambda + 2u_1''f_{\lambda\mu}f_{\mu\mu} + u_1'f_{\lambda\mu\mu}$. A positive sign of this expression would imply that an increase in risk will induce an increase in the investment in human capital. The converse is true if the derivative is negative. In the special case where $f(\lambda, \mu) = \mu\phi(\lambda)$, the sign of the second-order derivative will be positive (negative) depending only upon whether $u'''(c_1)c_1 + 2u''(c_1)$ is positive or negative.

[22] For a multiperiod analysis of a joint portfolio and savings choice model, see Levhari and T. N. Srinivasan.

[23] Consider, for instance, the following simple case. Let there be two consumption periods and let the market rate of interest be zero. Suppose the individual is offered the choice between two occupations with the following sequence of income distributions: $(\bar{y}, y)$ and $(y, \bar{y})$ where y is a random variable and its mean is $\bar{y}$. The distribution of present value $\bar{y} + y$ is of course identical in the two occupations but the individual will prefer the occupation in which uncertainty occurs *earlier in time*, basically because he has more information on which he may base his consumption decisions. The difference between the value of the two occupations is given by:

$$E\{\max_{c_0} [u(c_0, \bar{y} + y - c_0)]\} - \max_{c_0} [E\{u(c_0, y + y - c_0)\}] \geqq 0$$

See Jacob Marschak, pp. 200–05, and Drèze and Modigliani. If the utility function is strictly concave, the inequality will be a strict one.

$$\left.\frac{d\lambda^*}{dh}\right|_{h=0} = \frac{V_{c_0c_0}(-V_{\lambda h}) - V_{c_0\lambda}(-V_{c_0h})}{|\Delta|}$$

where

$$\Delta = V_{\lambda\lambda}V_{c_0c_0} - V_{c_0\lambda}^2 > 0$$

$$V_{c_0h} = -E\{u_1''(c_1)\phi(\lambda^*)(\mu-1)(1+r)\}$$

$$V_{\lambda h} = E\{u_1''(c_1)\phi(\lambda^*)(\mu-1)[-y_0(1+r) + \mu\phi'(\lambda^*)] + u_1'(c_1)\phi'(\lambda^*)(\mu-1)\}$$

$$V_{c_0c_0} = u_0''(c_0) + (1+r)^2E\{u_1''(c_1)\}$$

$$V_{c_0\lambda} = -(1+r)E\{u_1''(c_1)(-y_0(1+r) + \mu\phi'(\lambda^*))\}$$

The sign of $d\lambda/dh|_{h=0}$ is that of the numerator which assumes the form

$$\begin{aligned}&-u_0''(c_0)V_{\lambda h}\\ &-(1+r)^2E\{u_1''(c_1)\}E\{u_1'(c_1)\phi'(\lambda^*)(\mu-1)\}\\ &-(1+r)^2E\{u_1''(c_1)\}E\{u''(c_1)(-y_0(1+r)\\ &+\mu\phi'(\lambda^*))\phi(\lambda^*)(\mu-1)\}\\ &+(1+r)^2E\{u_1''(c_1)(-(1+r)y_0\\ &+\mu\phi'(\lambda^*))\}E\{u''(c_1)\phi(\lambda^*)(\mu-1)\}\end{aligned}$$

We wish to show that this expression is negative. Later it will be shown that $V_{\lambda h}<0$. Due to risk aversion

$$\phi'(\lambda^*)E\{u_1'(c_1)(\mu-1)\} = \phi'(\lambda^*)\ \mathrm{cov}\ \{u'(c_1);\mu\} < 0$$

It is thus sufficient to show that

$$\begin{aligned}&E\{u_1''(c_1)\}E\{u_1''(c_1)(-y(1+r) + \mu\phi'(\lambda^*))\phi(\lambda^*)(\mu-1)\}\\ &\geq E\{u_1''(c_1)(-(1+r)y_0 + \mu\phi'(\lambda^*))\}E\{u''(c_1)\phi(\lambda^*)(\mu-1)\}\end{aligned}$$

or after cancellations

$$E\{u_1''(c_1)\}E\{u_1''(c_1)\mu^2\} \geq [E\{u_1''(c_1)\mu\}]^2$$

This last inequality holds as a special case of Cauchy's inequality (see G. Hardy, J. Littlewood, and G. Polya, p. 16).

It remains to be shown that $V_{\lambda h}<0$. For this purpose it is sufficient to prove that

$$E\{u_1''(c_1)(\mu-1)[-y_0(1+r)+\mu\phi'(\lambda^*)]\}<0$$

Define μ' by $f_\lambda(\lambda^*, \mu')=1+r$. It follows from equation (7) in the text that $\mu'<E(\mu)$; that is, $\mu'<1$.

We may break the expectation into the following three integrals:

$$\begin{aligned}&\int_{\underline{\mu}}^{\mu'} u_1''(c_1)(\mu-1)[-y_0(1+r)+\mu\phi'(\lambda^*)]g(\mu)d\mu\\ &+\int_{\mu'}^{1} u_1''(c_1)(\mu-1)[-y_0(1+r)\\ &+\mu\phi'(\lambda^*)]g(\mu)d\mu\\ &+\int_{1}^{\bar{\mu}} u_1''(c_1)(\mu-1)[-y_0(1+r)\\ &+\mu\phi(\lambda^*)]g(\mu)d\mu\end{aligned}$$

The last integral is negative, and therefore it is sufficient to show that the *sum* of the first integrals is also negative.

For $\mu\in[\underline{\mu}, 1]$ the function $\mu_1''(c_1)(\mu-1)/\mu_1'(c_1)$ is monotone *decreasing* in μ; it follows that the sum of the first two integrals is smaller than

$$\frac{u''(c_1')(\mu-1)}{u'(c_1')}\int_{\underline{\mu}}^{1} u_1'(c_1)\cdot[-y_0(1+r)+\mu\phi(\lambda^*)]g(\mu)d\mu$$

For $\mu\in[1, \bar{\mu}]$, $\mu\phi'(\lambda^*)>y_0(1+r)$. Hence the sum of the two first integrals is also smaller than

$$\frac{u''(c_1')(\mu'-1)}{u^1(c_1')}\int_{\underline{\mu}}^{\bar{\mu}} u_1'(c_1)\cdot[-y_0(1+r)+\mu\phi'(\lambda^*)]g(\mu)d\mu$$

But from the first-order condition this last expression is zero.

REFERENCES

K. Arrow, *Essays in The Theory of Risk Bearing*, Amsterdam 1970.

G. Becker, *Human Capital*, New York 1964.

———, "Human Capital and The Personal Distribution of Income: An Analytical Approach," Inst. Publ. Admin., Univ. Michigan 1967.

Y. Ben-Porath, "Production of Human Capital and The Life Cycle of Earnings," *J. Polit. Econ.*, Aug. 1967, *75*, 352–65.

S. Bowles, "Schooling and Inequality from Generation to Generation," *J. Polit. Econ.*, May/June 1972, Part II, *80*, S219–51.

J. Drèze and F. Modigliani, "Consumption Decisions Under Uncertainty," *J. Econ. Theory*, Dec. 1972, *5*, 308–35.

G. Hardy, J. Littlewood, and G. Polya, *Inequalities*, Cambridge, Mass. 1967.

J. C. Hause, "Earnings Profile: Ability and Schooling," *J. Polit. Econ.*, May/June 1972, Part II, *30*, S108–38.

H. Houthakker, "Education and Income," *Rev. Econ. Statist.*, Feb. 1959, *41*, 24–28.

H. Leland, "Saving and Uncertainty: The Precautionary Demand for Savings," *Quart. J. Econ.*, 1968, *82*, 509–36.

D. Levhari and T. N. Srinivasan, "Optimal Savings Under Uncertainty," *Rev. Econ. Stud.*, Aug. 1969, *36*, 153–64.

J. Marschak, "Towards an Economic Theory of Organization and Information," in *Decision Processes*, New York 1954.

M. Nerlove, "On Tuition and the Cost of Higher Education: Prolegomena to a Conceptual Framework," *J. Polit. Econ.*, May/June 1972, Part II, *80*, S178–218.

J. Pratt, "Risk Aversion in The Small and in The Large," *Econometrica*, Jan.-Apr. 1964, *32*, 122–36.

M. Rothschild and J. Stiglitz, "Increasing Risk II," *J. Econ. Theory*, Mar. 1971, *3*, 68–84.

A. Sandmo, "The Effect of Uncertainty on Saving Decisions," *Rev. Econ. Stud.*, July 1970, *37*, 353–60.

T. Schultz, *Investment in Human Capital*, New York 1971.

———, *Human Resources*, New York 1972.

H. Shaffer, "Investment in Human Capital: Comment," *Amer. Econ. Rev.*, Dec. 1961, *51*, 1026–35.

L. Thurow, *Investment In Human Capital*, Belmont 1970.

B. Weisbrod, "Education and Investment in Human Capital," *J. Polit. Econ.*, Oct. 1962, *70*, 106–23.

Y. Weiss, (1971a) "Ability and Investment in Schooling: A Theoretical Note," *J. Econ. Lit.*, June 1971, *19*, 459–61.

———, (1971b) "Learning by Doing and Occupational Specialization," *J. Econ. Theory*, June 1971, *3*, 189–98.

———, "The Risk Element in Educational and Occupational Specialization," *J. Polit. Econ.*, Nov./Dec. 1972, *80*, 1203–13.

The Measurement of Technical Change Biases with Many Factors of Production

By HANS P. BINSWANGER*

Technical change biases have generally been measured in two-factor models, using value-added functions (see Paul David and Th. van de Klundert, Ryuzo Sato, and Theodore Lianos). A many-factor generalization of these procedures has the following advantages: Primary factors can be disaggregated into different classes of labor and capital, allowing an investigation of how technical change affects each of the subclasses. Intermediate inputs can also be included so that questions of natural resource biases or biases with respect to energy can be considered.[1] Also disaggregation is necessary when production processes are not separable between primary and intermediate inputs, which is a key assumption for fitting value-added functions.

The concept of Hicks neutrality is used in this paper. But it is used in a slightly amended version which leads to a definition of biases in terms of factor shares.[2]

$$B_i = \frac{d\alpha_i^*}{dt}\,\frac{1}{\alpha_i} \tag{1}$$

where α_i is the share of factor i. The notation $d\alpha^*$ indicates that relative factor prices are held constant for this share change. Technical change is i-saving if $B_i<0$, neutral if $B_i=0$, and i-using if $B_i>0$. This definition has the advantage that it leads to a single measure of bias for each factor in the n-factor case while Hicks' definition would lead to $n-1$ measures of bias for each factor.

Statistical estimates of biases and data on factor prices can be used to test the induced innovation hypothesis in the following way: Suppose innovation possibilities are neutral and factor prices are exogenous to the industry. Then a measured factor-saving bias should be associated with a rising factor price and vice versa. Further-

* Agricultural Development Council, New York and department of agricultural and applied economics, University of Minnesota. I wish to thank Richard A. King, Vernon W. Ruttan, James A. Seagraves, and T. Dudley Wallace for their helpful suggestions on earlier drafts. Research for this paper was supported by the U.S. Agency for International Development (USAID) through grants to the department of economics, North Carolina State University and to the Economic Development Center, University of Minnesota. The conclusions do not necessarily reflect the position of the USAID. This paper will be listed as Minnesota Agricultural Experiment Station Scientific Journal Series No. 8469.

[1] The author is presently engaged in such a study with Australian data.

[2] Hicks' definition is as follows: Technical change is said to be neutral, labor-saving, or labor-using depending on whether, at a constant capital-labor ratio, the marginal rate of substitution stays constant, increases, or decreases. Mathematically this can be expressed as follows:

$$\frac{d}{dt}MRS = \frac{d}{dt}\frac{f_K}{f_L} = -\frac{d}{dt}\frac{dL}{dK} = 0$$

where f_K and f_L stand for the marginal products and the capital-labor ratio is held constant. Neutrality is, therefore, a homothetic inwards shift of the unit isoquant. If at a constant factor ratio the marginal rate of substitution (or the ratio of the capital price to the labor price) is rising, then the labor share is declining. This leads immediately to definition (1). To estimate biases it is, however, not possible simply to look at historical factor share changes. The observed share changes have come about through biased technical change and through ordinary factor substitution in response to changes in the prices of the factors. The problem is to sort out to what extent the share changes have been due to biased technical change and to what extent to price changes. This can only be done, in a graphic sense, if the curvature of the isoquant is known. The substitution parameters of the production process have to be estimated before any biases can be measured.

more, turning points in trends of factor prices should be followed after some years by corresponding changes in the rates of biases. If, on the other hand, innovation possibilities are not neutral, then it is possible that a factor-using bias is associated with a rise in the price of the corresponding factor. Indeed, such a behavior proves the presence of biased innovation possibilities. All induced innovation can do is to offset the fundamental bias to some extent. But even in this case an acceleration of the price rise should result after some years in a decrease of the rate of the factor-using bias. The measured series of biases and factor prices will be inspected to check both for bias in invention possibilities and the sequence of turning points in factor prices and biases. For a detailed discussion of the induced innovation framework underlying this test, see the author (1974b).

Two models for measuring biases are considered, both using the translog cost function (see Laurits Christensen et al. (1970)).[3] Model A assumes variable rates of biases and is used to derive long-term series of biases. Model B assumes the biases to occur at constant rates and can be used with regression models. Since a cost function is used, very few constraints have to be imposed on the production process in both models (for example, it does not have to be homogenous of degree one).

Both models are applied to $U.S.$ agricultural data. For the period when data were available for both methods, the resulting estimates of bias were essentially the same, which gives support to the methodologies used.

I. The Translog Case

Model A

Every production function has a minimum cost function as its dual which relates factor prices to the cost of the output. The cost function contains all the information about the production process which the production function contains.

A minimum cost function with technical change in factor augmenting form and neutral economies or diseconomies of scale can be written

$$(2) \quad C = h(Y)\phi\left(\frac{W_1}{A_1}, \frac{W_2}{A_2}, \ldots, \frac{W_n}{A_n}\right)$$

where C is cost, Y is the output level, and W are the factor prices. The A are augmentation parameters corresponding to those in the dual production function. A proportional change in A_i has the opposite effect on cost of a proportional change in the price of factor i.

Let $R_i = (W_i/A_i)$ the factor price of the augmented factor unit (A_iX_i). The translog cost function can be written in logarithmic form as

$$(3) \quad \ln C = \ln\left[h(Y)\right] + \ln \nu_0 + \sum_i \nu_i \ln R_i + \frac{1}{2}\sum_i \sum_j \gamma_{ij} \ln R_i \ln R_j$$

where $h(Y)$ is a scale function of output and ν_0, ν_i, and γ_{ij} are the parameters of the cost function. The part before the double summation is a Cobb-Douglas function. If the cost function is Cobb-Douglas, then the production function will also be Cobb-Douglas (for proof, see Giora Hanoch).[4] We therefore can think of the terms in the double summation as amendments to the Cobb-Douglas function which change the elasticities of substitution away from unity (see equations (25) to (28) below). The function allows arbitrary and variable elasticities of substitution among factors.

[3] A general model for measuring biases in the many factor case for arbitrary twice differentiable production functions can be found in my doctoral dissertation.

[4] The parameters of the function are estimated later in the paper and the Cobb-Douglas restriction ($\gamma_{ij}=0$ for all i, j) is tested and rejected.

Equation (3) can be considered a functional form in its own right or can be regarded as a logarithmic Taylor series expansion to the second term around input prices of an arbitrary twice differentiable cost function (see Christensen et al. (1970)). With the proper set of constraints on its parameters, it can therefore be used to approximate any one of the known cost and production functions. The following symmetry constraint holds for all translog functions (equality of the cross derivatives).

$$(4) \qquad \gamma_{ij} = \gamma_{ji} \qquad \text{for all } i, j,\ i \neq j$$

The function must satisfy the following conditions:

i) *Linear homogeneity in prices:* When all factor prices double, the total cost has to double. It can be shown that this implies

$$(5) \qquad \sum_i \nu_i = 1; \quad \sum_i \gamma_{ij} = 0; \quad \sum_j \gamma_{ij} = 0 \qquad \text{for all } i, j$$

ii) *Monotonicity:* The function must be an increasing function of the input prices, i.e.,

$$(6) \qquad \frac{\partial\, ln\, C}{\partial\, ln\, R_i} = \nu_i + \sum_j \gamma_{ij}\, ln\, R_j \geq 0 \qquad i = 1, \ldots, n$$

iii) *Concavity in input prices:* This implies that the matrix

$$(7) \qquad \frac{\partial^2 C}{\partial R_i \partial R_j}$$

must be negative semidefinite within the range of input prices.[5]

To measure biases we need equations which explain factor shares in terms of factor prices. In augmented units, Shephard's lemma $\partial C/\partial W_i = X_i$ becomes

[5] This condition can be translated into the condition that the matrix (σ_{ij}) of partial elasticities of substitution be negative semidefinite. For the values of σ_{ij} found in the cross-sectional estimation, this was the case.

$$(8) \qquad \frac{\partial C}{\partial R_i} = \frac{\partial C}{\partial W_i}\frac{dW_i}{dR_i} = A_i X_i$$

The first derivatives of the translog function, with respect to the *log* of the factor prices, are equal to the shares:

$$(9) \qquad \frac{\partial\, ln\, C}{\partial\, ln\, R_i} = \frac{\partial C}{\partial R_i}\frac{R_i}{C} = \frac{(A_i X_i) R_i}{C} = \frac{A_i X_i (W_i/A_i)}{C} = \frac{W_i X_i}{C} = \alpha_i$$

Taking these derivatives, we have

$$(10) \qquad \alpha_i = \nu_i + \sum_j \gamma_{ij}\, ln\, R_j \quad i = 1, \ldots, n$$

Differentiating (10) totally

$$(11) \qquad d\alpha_i = \sum_{j=1}^{n} \gamma_{ij} d\, ln\, R_j \quad i = 1, \ldots, n$$

The proportional (*log*) change of a ratio is the difference of the proportional changes of its numerator and denominator. Then

$$(12) \qquad d\alpha_i = \sum_{j=1}^{n} \gamma_{ij}(d\, ln\, W_j - d\, ln\, A_j) \qquad i = 1, \ldots, n$$

Separating terms and using matrices

$$\begin{bmatrix} d\alpha_1 \\ \vdots \\ d\alpha_n \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} \\ \vdots & & \vdots \\ \gamma_{n1} & & \gamma_{nn} \end{bmatrix} \begin{bmatrix} d\, ln\, W_1 \\ \vdots \\ d\, ln\, W_n \end{bmatrix} - \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} \\ \vdots & & \vdots \\ \gamma_{n1} & & \gamma_{nn} \end{bmatrix} \begin{bmatrix} d\, ln\, A_1 \\ \vdots \\ d\, ln\, A_n \end{bmatrix}$$

or

$$(13) \qquad d\alpha = \gamma(d\, ln\, W) - \gamma(d\, ln\, A)$$

The matrix γ is not of full rank due to the homogeneity constraint. But calling an arbitrary factor the nth factor

$$(14) \qquad \gamma_{in} = -\sum_{i=1}^{n-1} \gamma_{ij}$$

Using (14) to remove γ_{in} from (13), we have

$$(15)\qquad d\alpha_i = \sum_{j=1}^{n-1} \gamma_{ij} dw_j - \sum_{j=1}^{n-1} \gamma_{ij} da_j$$

where $dw_j = dln\, W_j - dln\, W_n = dln(W_j/W_n)$ and $da_j = dln\, A_j - dln\, A_n = dln\, (A_j/A_n)$.

Let Γ be the truncated $(n-1)\mathrm{x}(n-1)$ matrix of the γ_{ij} which is of full rank. Then

$$(16)\qquad d\alpha_{(n-1)\mathrm{x}1} = \Gamma dw - \Gamma da$$

which gives us the solution for the changes in the A ratios

$$(17)\qquad da = dw - \Gamma^{-1} d\alpha$$

With the discrete time equivalent of (17), time-series of the augmentation series can be estimated, provided reliable estimates of the Γ^{-1} matrix are available. Going one step further, the share changes which would have occurred in the absence of factor price changes can be estimated directly. They are the share changes needed to estimate the biases according to equation (1). Call these changes $d\alpha^*$, which can be obtained from system (16) by setting $dw = 0$. Then

$$(18)\qquad d\alpha^* = -\Gamma da$$

Substituting da from (17),

$$(19)\qquad d\alpha^* = d\alpha - \Gamma dw$$

According to (19), we can immediately judge the nature of technical change for factors $i = 1, \ldots, n-1$.[6] To find what the factor share changes would have been had factor prices remained constant, simply subtract from the observed factor share changes that part which was caused by changing factor price ratios. The Γ matrix contains the substitution parameters which determine by how much the changes in factor price ratios alone could have altered the shares.

Before (19) can be used with time-series data, estimation of the coefficients of the matrix is necessary. This must be done with cross-section data where, ideally, all units are on exactly the same production function. We can then assume that all A_i are equal to one for all cross-sectional units and rewrite equation (10):

$$(20)\qquad \alpha_i = \nu_i + \sum_j \gamma_{ij}\, ln\, W_j + \epsilon_i \qquad i = 1, \ldots, n$$

Use this system of equations to estimate the γ_{ij} coefficients. Of course, one will never find a cross-section where all units are on exactly the same production function. Ways to deal with this problem are discussed in Section II and footnote 9.

Model B

Model A assumes that the rate of biases is not constant over time. For shorter time periods it is, however, possible to assume that the biases are constant. If this is done, biased technical change at constant exogenous rates can be introduced in the translog cost function in a similar way to that which Christensen et al. (1970) introduced it into the corresponding production function:

$$\begin{aligned}(21)\qquad ln\, C = {} & ln\, [h(Y)] + ln\, \nu_0 + \sum_i \nu_i\, ln\, W_i \\ & + \frac{1}{2} \sum_i \sum_j \gamma_{ij}\, ln\, W_i\, ln\, W_j \\ & + \nu_t\, ln\, \mathrm{t} + \omega_t (ln\, \mathrm{t})^2 \\ & + \sum_i \omega_i\, ln\, W_i\, ln\, \mathrm{t}\end{aligned}$$

where t stands for time.

Upon differentiation the share equations become

$$(22)\qquad \frac{\partial\, ln\, C}{\partial\, ln\, W_i} = \alpha_i = \nu_i + \sum_j \gamma_{ij}\, ln\, W_j + \omega_i\, ln\, \mathrm{t} \qquad i = 1, \ldots, n$$

[6] Since $\sum_{i=1}^{n} d\alpha_i^* = 0$, the solution for the nth factor is simply $d\alpha_n^* = -\sum_{i=1}^{n-1} d\alpha_i^*$.

which is estimation equation (20) with time entering as a variable, while ω_i is the constant exogenous rate of the bias of factor i.

If (22) is used as a regression equation with a time-series or a combination of cross-section and time-series, the introduction of time in this way will ensure that biased technical change at constant rates will not bias the econometric estimates of γ_{ij}. Furthermore, the coefficients ω_i can be used to derive another set of price corrected shares series, say $d\alpha_i^{**}$, which can be used with equation (1) to estimate the biases for the particular period

$$(23) \qquad d\alpha_i^{**} = \hat{\omega}_i d\, ln\, t \qquad i = 1, \ldots, n$$

Of course, this model cannot be used to extrapolate outside of the short regression period because then the assumption of a constant exogenous rate of bias is tenuous.

II. Cross-Sectional Estimation of the Parameters of the Cost Functions

The cross-sectional estimation of the cost function used state data for the United States. A more detailed account of this estimation and the data used are given in my (1974a) article.

Four sets of cross-section data were obtained for thirty-nine states or groups of states. The cross-sections were derived from census and other agricultural statistics for the years 1949, 1954, 1959, and 1964.

In general, Zvi Griliches' (1964) definitions of factors were used. He distinguishes the following five factors: land, labor, machinery, fertilizer, and all others. Intermediate inputs are included in this list and the function fitted corresponds to a gross-output function rather than a value-added function.[7]

The estimation equations are[8]

$$(24) \quad \alpha_{ikt} = \nu_i + \sum_j \gamma_{ij}\, ln\, W_{jkt} + \omega_i\, ln\, t + \sum_r \delta_{ir} d_r + \epsilon_{ikt}$$

where:

$i, j = 1, \ldots, 4$ are the indices for the factors of production

$k = 1, \ldots, 39$ are the indices for the states

$t = 1, \ldots, 4$ are the indices for the time periods

δ_{ir} = share specific regional dummies for $r = 1, \ldots, 4$

$d_r = 1$ if $k \in r$; 0 if $k \notin r$

The estimation equations are unaltered if neutral efficiency differences exist among states or time periods. Neutral differences would only alter the intercept ν_0 of the cost function which drops out upon differentiation. If any omitted factor, such as education or research and extension, affects efficiency neutrally, leaving it (them) out of the estimation equation will not bias the results.[9]

---

[7] Most of the cross-section data come from published *USDA* sources (see the author (1973)). Expenditures on factors usually are actual expenditures, and where applicable, imputed expenditures for wages of family members, interest charges, depreciation, and taxes. Quantity data are derived as price weighted indexes of physical units (land and fertilizer), or the sum of individually deflated expenditures (all other), or a combination of these methods (machinery and labor). The quantity data were already computed by Gideon Fishelson, who used Griliches' (1964) data with slight changes. Expenditure and quantity data are consistent with each other. The price data were obtained by dividing the expenditure data by the quantities.

[8] Due to the homogeneity constraint (5), only $n-1$ share equations are linearly independent and can be estimated simultaneously.

[9] That equation (24) measures the γ_{ij} parameters of the functional form (3) can be seen as follows: If there is only one cross-section, and all cross-sectional units have the same A_i parameters (no efficiency differences), factors can be rescaled to make all A_i equal to one. Hence $W_{jk} = R_{jk}$. If there are neutral efficiency differences among the cross-sectional units, all the A_i will differ by the same proportion among the two units. This proportion can be absorbed in a separate intercept term for each unit. These intercepts drop out upon differentiation so that we can again set all $A_i = 1$. If more than one cross-section is used, (24) still gives the correct estimates for (3), provided that for all cross-sectional

Nonneutral efficiency differences among the observational units will have the effect that the true α_i will differ for each observational unit; at equal factor prices, shares will not be equal.[10] If such differences occur among all units, the estimates of the coefficients of (24) will be biased. However, if such differences occur only among groups of states, the proper set of regional dummies will again lead to unbiased estimators. Regional dummies distinguishing five regions were therefore included in the regression equations. Nonneutral differences might arise due to educational differences, differences in research and extension, or differences in product mix.

If (24) is estimated with time-series data, a time trend in the estimation equation will solve the problem of estimation of biases over time, as explained in Model *B*, provided the rates of biases stayed reasonably constant during the estimation period.

A detailed account of the error specification problem is given in my (1974a) article. Problems arise from the fact that 1) time-series and cross-section data are combined and 2) that within each cross-section, the error terms of the shares equations are not independent.[11]

Both problems could not be handled simultaneously. For estimation purposes all cross-sections were combined and restricted generalized least squares (see Theil) applied to the four share equations, as if there were no problem of error interdependence over time. While this leads to consistent estimates of the γ_{ij} coefficients, there is an efficiency loss and *t*-ratios of the estimates will be overstated to some extent.[12]

Prior to the estimation, several constraints were tested by applying the model to the four cross-sections individually using restricted generalized least squares (*GLS*). These tests, therefore, did have the desired asymptotic properties. The homogeneity constraint was not rejected in any of the four cross-sections (.05 significance level). The symmetry constraint was rejected only in the 1964 data set. However, the Cobb-Douglas constraint ($\gamma_{ij}=0$ for all i, j) was rejected in all cross-sections.[13] Hence, homogeneity and symmetry were imposed in the estimation with all data pooled.

Pooling the cross-sections implies constancy of the γ_{ij} coefficients over time. This was tested as follows: a two-equation *GLS* model is fitted for each share with the 1949 data used for the first equation and the 1959 data for the second equation. The homogeneity constraint was imposed on the data.

The hypothesis is never rejected at the .01 level of significance, although it is rejected in two equations at the .05 level

units each A_i changes at the constant rate ω_i during the period under investigation.

[10] Nonneutral efficiency differences between two states implies that at equal factor prices, the states will use factors in differing proportions. Factor shares will, therefore, differ even at equal factor prices.

[11] *Problem* 1): For each share equation, data from four cross-sections are combined. This poses the familiar problem of combining cross-section and time-series data. Despite the 5-year interval between the cross-sections, this problem is still important: the correlation coefficients of the ordinary least squares (*OLS*) residuals of the share equations between the cross-sections of 1949 and 1965 ranged between 0.62 and 0.87. If first-order autocorrelation was the true error specification over time, this would imply first-order autocorrelation coefficients larger than 0.9. *Problem* 2): Within each of the 4 cross-sections, the error terms of the $n-1$ estimation equations are not independent, since for each state the same variables which might affect the shares in addition to the prices were left out of the model. If restrictions across equations ($\gamma_{ij}=\gamma_{ji}$) are imposed, *OLS* estimators are no longer efficient, despite the fact that all equations contain the same explanatory variables on the right-hand side (see Henri Theil). Therefore, the seemingly unrelated regression problem applies.

[12] The computer program used was Triangle Universities Computing Center.

[13] All three tests were performed against a completely unconstrained model.

TABLE 1—RESTRICTED ESTIMATES OF THE COEFFICIENTS OF THE TRANSLOG COST FUNCTION AND t-RATIOS[a,b,e]

	Independent Variables										
Factor	Land	Labor	Machinery	Fertilizer	Other[c]	Year	Intercept	*MN*	*GF*	*SE*	*GS*[d]
Land	.07747	−.03613	.00478	.01066	−.05678	.00847	.2603	−.1021	−.0394	−.1073	−.0577
	(6.02)	(3.25)	(.47)	(2.14)		(1.47)	(9.96)	(10.2)	(4.1)	(8.9)	(4.7)
Labor		−.06367	−.00661	−.02805	.13446	−.05482	.5218	.0194	−.0016	.0169	.0246
		(3.67)	(.59)	(4.97)		(9.08)	(14.91)	(1.63)	(.15)	(1.09)	(1.63)
Machinery			−.03485	−.00877	.04545	.02498	.0926	−.0033	.0369	−.0186	.0072
	Symmetric		(1.31)	(.97)		(4.66)	(3.46)	(.41)	(5.08)	(1.86)	(.73)
Fertilizer				.00068	.02548	.00178	.0745	.0104	−.0041	.0370	−.0024
				(.12)		(.63)	(5.6)	(2.5)	(1.10)	(7.24)	(.49)
Other					−.14861						

Source: USDA (see the author (1973)).

[a] Restrictions imposed: $\gamma_{ij}=\gamma_{ji}$ and $\sum_{j=1}^{n}\gamma_{ij}=0$ for all i, j.

[b] Critical values with 578 degrees of freedom are $t_{.05}=1.96$ and $t_{.01}=1.65$; t-ratios may be overstated due to error interdependence over time.

[c] Implied estimates computed using the homogeneity constraint.

[d] *MN, GF, SE, GS* are dummies for Mixed Northern agriculture, Grain Farming States, Southeast, and Gulf States, respectively. The intercept stands for Western States and the coefficients of *MN, GF, SE, GS* are deviations from this intercept.

[e] Since *GLS* methods were used, no R^2 are reported due to the ambiguous interpretation of these statistics in *GLS* models. For shares equations in individual years estimated by *OLS* methods, the R^2 ranged from 0.5 to 0.9.

of significance. A test for constancy of coefficients can be interpreted as a test of the factor augmenting hypothesis, if there are no other specification errors. This test, therefore, implies a weak support for this hypothesis.

Table 1 reports the estimates of the pooled regressions. The t-ratios of the price coefficients appear to be low, despite the fact that they may be overstated to some extent. However, $\gamma_{ij}=0$ implies that the corresponding partial elasticity of substitution is equal to one (see equations (25) and (26)). Therefore, we would expect, a priori, some of the γ_{ij} coefficients to be zero.

In the labor and the machinery equations, the coefficients of time are significant. They imply that technical change has been labor-saving and machinery-using. Significant coefficients of regional dummies imply nonneutral regional efficiency differences.

The estimates of the γ_{ij} coefficients can be converted into point estimates of Allen partial elasticities of substitution (σ_{ij}) and of elasticities and cross elasticities of factor demand (η_{ij}) according to the following equations (for proof see my (1974a) article):

$$(25)\quad \sigma_{ij} = \frac{\gamma_{ij}}{\alpha_i\alpha_j} + 1 \qquad \text{for all } i \neq j$$

$$(26)\quad \sigma_{ii} = \frac{1}{\alpha_i^2}(\gamma_{ii} + \alpha_i^2 - \alpha_i) \qquad \text{for all } i$$

$$(27)\quad \eta_{ij} = \frac{\gamma_{ij}}{\alpha_i} + \alpha \qquad \text{for all } i \neq j$$

$$(28)\quad \eta_{ii} = \frac{\gamma_{ii}}{\alpha_i} + \alpha_i - 1 \qquad \text{for all } i$$

The γ_{ij} coefficients have little intuitive meaning, therefore it is easier to evaluate them by what they imply for these elasticities. Table 2 shows the results using the unweighted average factor shares of the thirty-nine states in the period 1949–64.

The matrix of elasticities of substitution is negative semidefinite, which implies that the matrix of cross derivatives of the cost function is negative semidefinite, i.e., that the cost function is concave.[14] All own demand elasticities have the correct sign.

[14] The elasticities of substitution also show that the production process is not separable between primary and intermediate inputs. Separability would imply, *inter alia*, that the partial elasticities of substitution of fertilizer with the primary factors land, labor, and machinery be equal (see Ernst Berndt and Christensen). This is obviously not the case.

TABLE 2—ESTIMATES OF THE PARTIAL ELASTICITIES OF SUBSTITUTION AND OF FACTOR DEMAND WITH RESPECT TO OWN PRICE[a]

	Land	Labor	Machinery	Fertilizer	Other[b]
Elasticities of substitution					
Land	−2.225	.204	1.215	2.987	−.031
	(.57)	(.24)	(.46)	(.93)	
Labor		−3.028	.851	−1.622	2.224
		(.19)	(.25)	(.53)	
Machinery			−7.379	−.672	1.844
	Symmetric		(1.23)	(1.72)	
Fertilizer				−26.573	2.961
				(4.61)	
Other					−2.852
Elasticities of factor demand	−.336	−.911	−1.089	−.945	−1.042
(η_{ii})	(.09)	(.06)	(.18)	(.16)	

Source: See Table 1.
[a] Standard errors in parentheses are computed as follows:

$$SE(\sigma_{ij}) = \frac{SE(\gamma_{ij})}{\alpha_i\alpha_j}; \quad SE(\eta_{ii}) = \frac{SE(\gamma_{ii})}{\alpha_i}$$

[b] Implied estimates computed using the homogeneity constraint.

The demand for land appears very inelastic. The demand elasticities for machinery and other inputs are larger than 1, a fact to keep in mind, since it implies that a rise in the corresponding factor prices will, other things being equal, lead to a fall in the factor share.

III. The Empirical Measures of Biases in Efficiency Gains

This section presents the derived series of biases for the years 1912 to 1968 using Model *A*. It also presents the series of actual factor shares and factor prices. The data come from published *USDA* sources. The variables are constructed so that they correspond as closely as possible to the variables used in the cross-section analysis. Total correspondence was, however, not achievable.[15] For a detailed account of the data, see my doctoral dissertation.

The basic estimation equation for the biases are equations (19) which when expanded reads

$$d\overset{*}{\alpha}_i = d\alpha_i - \sum_{j=1}^{n-1} \hat{\gamma}_{ij} dw_j \qquad i = 1, \ldots, n-1 \tag{19a}$$

where the $d\alpha_i^*$ is the change in the share of factor i in the absence of ordinary factor substitution due to price changes; $d\alpha_i$ is the actual total change in share i, which includes the effect of the price changes; dw_j is the proportional change of the ratio of the price of factor i to the price of other inputs. For actual estimation purposes, series of three-year moving averages of the shares and the factor prices were constructed. Then discrete differences of these moving averages at four-year intervals were taken and used in the discrete change equivalent of (19). The γ_{ij} coefficients are taken from Table 1. The resulting $d\alpha_i^*$ can be substituted into the discrete equivalent of equation (1) to compute rates of biases. Here, however, we compute series α_i^* which show how the shares would have developed after 1912 in the absence of factor price changes:

$$\overset{*}{\alpha}_{it} = \alpha_{i,1912} + \sum_{\tau=0}^{t} \Delta\overset{*}{\alpha}_{it} \tag{29}$$

[15] No quality adjustments were made for any of the factors because we are really not interested in earnings and biases of some adjusted efficiency unit of labor or land, but in the earnings of natural units of factors. Any quality adjustment before measuring the biases is, therefore, inappropriate.

These series are presented in Panel *A* of Table 3. Series of standardized values, i.e., $R_{it}=\alpha^*_{it}/\alpha_{i,1912}$ are presented in Figure 1, which is in semilogarithmic scale. Hence, the slopes of the lines indicate biases according to equation (1), while the position of the line shows the cumulative bias since 1912.

How good are these series? From its assumptions the approach is not very restrictive. No separability is imposed on the production process, an assumption which is implicit whenever a production function is used with only capital and labor. This cost function approach also allows nonconstant returns to scale, but requires that the dual production function be homothetic. The model assumes simple cost minimization. This assumption is not violated even if the government intervenes in factor and goods markets. Government intervention would be recorded correctly either in quantities or price series and not bias the results. Only government control of both prices and quantities would violate the assumption.

The key assumption of the approach is the constancy of the γ_{ij} parameters over time. This assumption was not supported as well as one might wish, but it cannot be avoided since data to estimate γ_{ij} for the beginning of the period are not available. The quality of the α^* series depends completely on the quality of γ_{ij} estimates. If they were really totally wrong, chances would be that over the long periods involved, which include two World Wars and the Great Depression, some strange result would be immediately apparent in the α^* series. Such a result might be if one of the α^* series became negative. Smaller errors in the γ_{ij} coefficients are, of course,

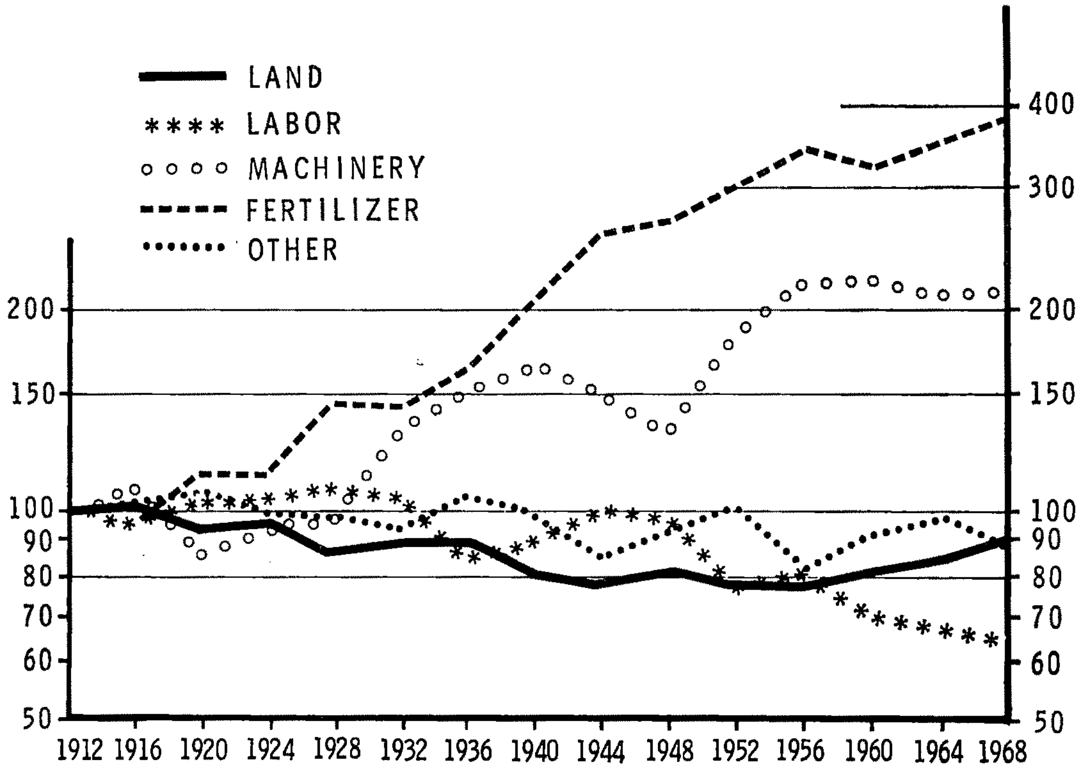


FIGURE 1. *U.S.* INDICES OF BIASES IN TECHNICAL CHANGE: MODEL *A* ESTIMATES OF $\alpha^*_{it}/\alpha_{i,1912}$; 1912=100

TABLE 3—PRICE CORRECTED SHARES (α_i^*), ACTUAL FACTOR SHARES, AND FACTOR PRICES USED IN COMPUTING (α_i^*)

	Year	Land	Labor	Machinery	Fertilizer	Other
A: Price Corrected Factor Shares α_i^* Model *A* Estimates	1912	21.0	38.3	10.9	1.9	28.0
	1916	21.2	36.7	11.6	1.8	28.7
	1920	19.6	39.3	9.3	2.1	29.7
	1924	20.0	39.7	10.3	2.2	27.8
	1928	18.1	41.4	10.4	2.7	27.4
	1932	18.8	40.3	14.3	2.7	24.0
	1936	18.9	32.5	16.3	3.0	29.3
	1940	16.8	34.3	17.6	3.9	27.5
	1944	16.5	38.4	16.1	4.8	24.2
	1948	17.1	37.2	13.9	5.1	26.7
	1952	16.5	29.8	19.7	5.7	28.3
	1956	16.3	30.6	23.1	6.5	23.4
	1960	17.1	27.2	23.4	6.1	26.1
	1964	17.8	25.8	22.4	6.7	27.3
	1968	19.1	25.3	23.1	7.2	25.3
B: Actual Factor Shares α_i	1912	21.0	38.3	10.9	1.9	28.0
	1916	21.6	36.5	11.6	1.9	28.4
	1920	17.3	40.5	10.1	2.0	30.1
	1924	19.7	38.5	10.3	1.7	29.7
	1928	15.9	40.9	10.2	1.9	31.1
	1932	18.6	37.6	12.6	1.6	29.7
	1936	14.9	34.7	14.5	2.2	33.7
	1940	12.0	35.3	15.1	2.3	35.2
	1944	8.5	39.5	14.0	2.3	35.6
	1948	9.4	37.7	12.2	2.4	38.3
	1952	9.8	29.7	17.5	3.0	40.0
	1956	11.5	27.4	20.1	3.3	37.8
	1960	15.6	21.3	19.8	2.9	40.4
	1964	17.5	18.3	18.5	3.3	42.3
	1968	20.4	15.8	19.1	3.6	41.1
C: Factor Prices Relative to Agricultural Output Prices 1912=100	1912	100.0	100.0	100.0	100.0	100.0
	1916	113.3	106.8	110.0	105.7	103.8
	1920	79.0	104.3	81.3	85.7	105.0
	1924	119.0	134.5	111.7	93.1	106.6
	1928	104.8	154.1	128.5	90.0	118.9
	1932	160.8	194.7	231.5	128.6	101.5
	1936	69.4	113.4	189.2	99.6	110.9
	1940	87.3	179.0	288.8	103.4	160.1
	1944	65.2	217.2	244.2	63.0	211.7
	1948	73.0	247.8	226.6	50.4	222.8
	1952	91.3	274.3	301.1	53.6	214.6
	1956	145.8	407.9	423.7	65.9	229.6
	1960	254.1	502.7	550.3	63.0	241.5
	1964	338.1	610.0	651.2	63.2	270.9
	1968	481.0	766.9	735.8	58.2	280.4

Source: See Table 1.

not ruled out by such considerations. The errors could even be large enough to make inferences from small direction changes of the series impossible. Some conclusion on the quality of the α^* series and the underlying γ_{ij} estimates can be obtained by comparing the Model A estimates presented above to the Model B estimates which use the time coefficient of the regressions reported in Table 1. The price corrected share changes $\Delta\alpha^{**}$ for Model B are computed for the period 1948–64 according to equation (23) under the assumption that the rate of the bias for each factor remained constant during that particular period or, alternatively, that ω_i parameters measure an average rate of bias. Apart from the fact that γ_{ij} and ω_i were estimated in the same equations, the Model A estimates $\Delta\alpha_i^*$ have nothing to do with Model B estimates $\Delta\alpha_i^{**}$; therefore there is no reason, apart from chance, that they would be similar if either set of estimates were incorrect. Table 4 shows the comparison of the Model B estimates with the Model A estimates reported in the graphs.

Both series estimate biases with the same sign and of about the same magnitude. Because of the differences in the underlying assumption, they cannot be expected to agree perfectly. The consistency of the estimates provides support for both methodologies of measuring biases

IV. Conclusions

Table 3 and Figure 1 show a very strong fertilizer-using bias, accompanied by a rapid decline of the price of fertilizer relative to the price of agricultural output. Over a long period of time it is safe to assume that the fertilizer price is governed by cost conditions in the fertilizer industry, i.e., that it is exogenous to agriculture. The behavior of the fertilizer bias is thus consistent with induced innovation.

It is also reasonable to assume that the price of labor is to a great extent governed by wage rates in the nonagricultural sector, and exogenous to agriculture to a substantial extent. The price of labor increased throughout the period, but the rise was much more rapid after about 1940 than before. Up to 1944 technical change was labor neutral, but thereafter a strong labor-saving bias did exist (which agrees with Lianos' findings). The behavior of price and bias is again consistent with the induced innovation hypothesis and neutral innovation possibilities. Between 1944 and 1968 the observed labor share dropped from almost 39.5 to 15.8 percent, i.e., a drop of 23.7 percent. Biased technical change alone could have explained a drop of almost 14 percent. This

TABLE 4—COMPARISON OF MODEL A AND B ESTIMATES OF BIASES FOR THE PERIOD 1948–64 FOR THE UNITED STATES

Factor	1948 Level of Shares[a]	Estimated Share Change Due to Technical Change Alone[a] Model A 1948–64 ($\Delta\alpha^*$)	Model B 1948–64 ($\Delta\alpha^{**}$)
Land	9.4	+ 2.3	+ .7
Labor	37.7	−15.1	−11.4
Machinery	12.2	+ 6.9	+ 8.5
Fertilizer	2.8	+ .5	+ 1.6

Source: See Table 1.
[a] Shown in percent.

means that about two-thirds of the drop in the labor share must be explained by biased technical change and only one-third by simple price substitution effects.

The price of machinery[16] in the long run is also primarily governed by cost conditions in the machinery producing and machinery repairing industries and by fuel costs, and therefore also regarded as exogenous to agriculture. The overall rise of machinery prices was about the same as in labor prices. For both factors, the rate of the price rise accelerated after 1948. But despite that price rise, technical change was machinery-using, not saving. Had innovation possibilities been neutral, this could not have occurred. Innovation possibilities must have been machinery-using regardless of the role of factor prices in determining biases. Any price induced bias would have been machinery-saving, not machinery-using. If price induced biases are important, then the machinery-using bias would have been even larger in the absence of the rise in machinery prices.

Fertilizer and labor biases are consistent with the hypothesis of neutral innovation possibilities and the hypothesis that factor prices account for most of the biases, but the machinery bias contradicts this. It is even possible that the labor and fertilizer biases were primarily fundamental biases and that the corresponding price changes were coincidental rather than causative.

On the other hand, the view that all biases are caused by nonneutral innovation possibilities is contradicted by the time sequence of turning points in factor prices and biases. The labor price rise strongly accelerated at the start of World War II; and it is between six to ten years later that a strong labor-saving bias emerges. Similarly, the machinery price really takes off between 1948 and 1952 and it is again about six years later that the machinery-using bias disappears. Robert E. Evenson has found a mean lag between research initiation and benefits of between five and one-half and eight and one-half years in agriculture, which would support causality in the observations above. Of course, these changes might be coincidental, but I do not believe that this is very likely.

The series on land bias and price cannot give us much more information because the land price is largely endogenous to agriculture. Land prices fluctuated widely but there was apparently very little bias with respect to land. The price of other inputs rose especially fast between 1936 and 1944, but did not change much before or after. However, there was no discernible bias with respect to this conglomerate of factors.

The clearest conclusion to be drawn from the series is that fundamental biases in innovation possibilities were an important source of the machinery-using bias in *U.S.* agriculture.

In addition to that clear conclusion, there are indications that very large changes in factor prices can have a strong impact on the direction of technical change. The fertilizer price fell through most of the period relative to the price of the four other inputs. This fall was accompanied by a strong fertilizer-using bias. And the lags between turning points of price trends and biases for machinery and labor indicate price responsiveness of biases. But it is also clear from the series that it takes very substantial changes in factor prices in order to perceptibly influence the biases. The direction of technical change may respond only to massive changes in relative prices.

[16] The machinery price is the price of machinery services, which includes interest charges, depreciation, fuel, and repair costs. Depreciation rates and repair costs have strongly increased over time.

REFERENCES

E. R. Berndt and L. R. Christensen, "The Internal Structure of Functional Relationships: Separability, Substitution, and Aggregation," *Rev. Econ. Stud.*, forthcoming.

H. P. Binswanger, "The Measurement of Biased Efficiency Gains in U.S. and Japanese Agriculture to Test the Induced Innovation Hypothesis," unpublished doctoral dissertation, No. Carolina State Univ. 1973; Microfilms, Ann Arbor.

———, (1974a) "A Cost Function Approach to the Measurement of Factor Demand Elasticities and of Elasticities of Substitution," *Amer. J. Agric. Econ.*, May 1974, *56*, 377–86.

———, (1974b) "A Microeconomic Approach to Induced Innovation," *Econ. J.*, Dec. 1974.

L. R. Christensen, D. W. Jorgenson, and L. J. Lau, "Conjugate Duality and the Transcendental Logarithmic Production Function," unpublished paper presented at the Second World Congress of the Econometric Society, Cambridge, Sept. 1970.

———, ———, and ———, "Transcendental Logarithmic Production Frontiers," *Rev. Econ. Statist.*, Feb. 1973, *55*, 28–45.

P. A. David and Th. van de Klundert, "Biased Efficiency Growth and Capital-Labor Substitution in the U.S., 1899–1960," *Amer. Econ. Rev.*, June 1965, *55*, 357–94.

W. E. Diewert, "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," *J. Polit. Econ.*, May/June 1971, *79*, 481–505.

R. E. Evenson, "The Contribution of Agricultural Research and Extension to Agricultural Production," unpublished doctoral dissertation, Univ. Chicago 1968; Microfilms, Ann Arbor.

G. Fishelson, "Returns to Human and Research Capital, United States Agriculture, 1949–1964," unpublished doctoral dissertation, No. Carolina State Univ. 1968; Microfilms, Ann Arbor.

Z. Griliches, "The Demand for Fertilizer: An Economic Interpretation of a Technical Change," *J. Farm Econ.*, Aug. 1958, *40*, 591–606.

———, "Research Expenditure, Education, and the Aggregate Agricultural Production Function," *Amer. Econ. Rev.*, Dec. 1964, *54*, 961–74.

G. Hanoch, "Generation of New Production Functions through Duality," Harvard Inst. Econ. Res. disc. pap. 118, 1970.

T. P. Lianos, "The Relative Share of Labor in the United States Agriculture," *Amer. J. Agr. Econ.*, Aug. 1971, *53*, 411–22.

R. Sato, "The Estimation of Biased Technical Progress," *Int. Econ. Rev.*, June 1970, *11*, 179–207.

R. M. Solow, "Some Recent Developments in the Theory of Production," in M. Brown, ed., *The Theory and Empirical Analysis of Production*, Nat. Bur. Econ. Res. *Stud. Income and Wealth*, Vol. 31, New York 1967.

H. Theil, *Principles of Econometrics*, New York 1971.

Triangle Universities Computing Center, *Program Manual for Two and Three Stage Least Squares*, Research Triangle Park 1972.

Household Production Theory, Quality, and the "Hedonic Technique"

By John Muellbauer*

A feature of many of the empirical papers which use the "hedonic technique" of correcting prices for quality change is their reference to work on household production by Kelvin Lancaster and Richard Muth (and sometimes Gary Becker) as providing a theoretical foundation. However, the exact nature of this theoretical support has never been adequately spelled out.

Roughly speaking, exponents of the hedonic technique (see Zvi Griliches (1971) for a very useful survey and also Jack Triplett and Robert Gordon) take the view that the quality of a good is related to measurable specification variables or characteristics such as size, performance, etc. All the empirical applications regress prices or the *logs* of prices of the different varieties or models of a type of good on these characteristics. However, there are two main variants of the empirical forms which have been used: single year cross-section regressions, and pooled (over at least two years) time-series/cross-section regressions.

The first variant claims to estimate the "shadow prices" of the "characteristics" of the goods for a given year. These shadow prices are then used to price some average bundle of characteristics in the two years which are being compared. If, say, the base year's level of characteristics is used, an index which looks similar to a Laspeyres price index can be constructed. In the second method, the time dummy variables which link the cross-sections are supposed to pick up differences in general price levels at different times. The inclusion of the specification variables then controls for the effects of changes in quality, i.e., constrains quality to be a constant. In other words, the variation in the dependent variable (price or *log* of price) is broken up into two parts. One reflects changes in the average price level over time, the other reflects differences in characteristics.

The main purpose of this paper is to relate the hedonic technique to a theoretical constant utility price index when quality changes are taken into account. It shows that the conditions under which precise statements can be made about the relationship of the usual empirical approximations to such a theoretical price index are rather restrictive. These conditions impose interesting and quite strong restriction on the classes of empirical techniques of quality correction which can thus be rigorously justified.

The approach is unashamedly one-sided; only the demand side is treated. The orientation of the paper is towards the analysis of the welfare of individual consumers and less towards the explanation of market phenomena. Its subject matter is therefore rather different from that of the recent paper by Sherwin Rosen. The supply side and the simultaneity problems which may arise are ignored. This is done partly for reasons of space given that the analysis of demand is prior to market analysis. How-

* Lecturer in economics, Birkbeck College, University of London. I am grateful to Keith Cowling, Hugh Davies, Ben Fine, Zvi Griliches, Bertie Hines, Yehuda Kotovitz, and Robert Pollak for helpful discussions. Valuable comments were received from the referee and the managing editor. However, I take full responsibility for remaining errors and ambiguities.

ever, sometimes it may also be realistic, as with good second-hand markets, to assume that consumers are price takers and adjust fully.

The structure of the paper is as follows. After the summary and conclusions, which are hopefully self-contained, Section II spells out the basic utility maximizing model within the household production framework. A true (constant utility) price index is defined. Section III introduces quality change and discusses the difficulty of taking account of this in the true price index. Section IV discusses conditions under which these difficulties can be overcome. Section V discusses some special household production models, in particular that of Lancaster. Section VI introduces an alternative approach outside the household production framework. This is based on a generalization of the one-good model proposed by Fisher-Shell. Section VII deals with a quite different strand in the literature originating in Hendrik Houthakker's 1952 paper. Here shadow prices are assumed to exist as part of the supply conditions which are external to the household. One of the major conclusions, therefore, of the present paper is that there are *three* separate classes of theories that can be said to underly empirical work on quality change.

I. Summary and Conclusions

The development of the theory and the conclusions of the paper as regards the first of these approaches may be summarized as follows.

A true price index is defined as the relative expenditure under two price regimes required to reach a given level of utility. If tastes are constant and there is no quality change, the form of the functional relationship between expenditure on the one hand and prices and utility on the other is constant. In this case the traditional Laspeyres and Paasche approximations to a true index are valid and well known. However, if quality changes, the form of this functional relationship changes. It then follows that the traditional approximations are in general no longer valid. The solution that the household production approach then proposes is to hypothesize that a household derives utility from some basic goods (Z) whose qualities do not change and that these are produced from market inputs (x) whose qualities can change. If one can at least identify the shadow costs (π) of Z, one should be able to approximate the true price index without necessarily knowing the full household technology and tastes.

However, if the index is to make much sense, the shadow value of Z must have a stable relationship with the market value of x at a given technology. This "consistency condition" is satisfied if the technology is homogeneous of any positive degree in inputs and outputs. A Laspeyres index is then defined on π (and the prices of other market goods, if any).

In conventional theory without quality change, a Laspeyres index is an upper bound on the "true" index using base utility as reference. The reason why this may not be true in the presence of quality change is that the relative π which correspond to the actual current period purchases may be different from the relative π which are not directly observable and which correspond to current market prices and *base utility*. There are two alternative conditions which will guarantee that this problem does not arise. The first is a restriction on the technology (specified as a form of separability of the cost function for Z) which implies a linear transformation locus in Z and in addition, constant returns to scale in the production function relating Z to x. This restriction is equivalent to nonjointness. Nonjointness implies that the level of each input x_i can be split into separate parts, each of which is used

in only one branch of production, i.e., for only one Z_j. This special case is an example of Paul Samuelson's well-known "nonsubstitution theorem."

Nonjointness, however, is unrealistic for many examples of household production and poses serious problems in practical identification of Z. For example, the use of an automobile produces travel services and psychic rewards jointly. The theoretical advantage that consumers with different tastes and incomes will then face the same shadow costs (given the same technology for each consumer) is therefore bought at a high price. But this is the only case when the "aggregation problem" cannot arise. The theorems which I shall prove to establish this proposition imply a criticism of some of the work of the household production theorists. Much of the alleged usefulness of their approach stems from the ability to argue about likely consequences on the Z of changes in π. However, if different consumers face a different π, little information in the form of empirical restrictions on behavior is gained by making the usual a priori assumptions of the household production theorists.

The second alternative is to assume that the utility function is homothetic in Z. It then follows that, at given prices and under constant returns, the shadow costs will be the same regardless of the consumer's utility level. However, homotheticity which implies unitary income elasticities is not supported by empirical evidence.

These general problems can be illustrated by a discussion of some special cases. A good example is the Lancaster model, earlier proposed by W. M. Gorman, with, in addition, a variation such that Z is defined as the sum of characteristics with the *logs* of x instead of x as weights. For goods purchased by the consumer, the Lancaster model implies a linear relationship between the market prices and the characteristics with π as parameters. This can be interpreted as a justification of the first version of the empirical hedonic technique mentioned in the introduction. Gorman's variation implies that x enters the price-characteristics relationship. I prove a simple theorem which shows that a semilog price-characteristic relationship cannot be directly interpreted in the household production framework. However, a *partial* semilog relationship between the price and any *single* characteristic is permitted. This is an important empirical restriction, since the semilog form has been widely used in applications.

Moreover, the Lancaster model is not altogether plausible, particularly with regard to durable goods which are the most frequent objects of study in the empirical hedonic literature. For, although the divisibility assumption is just as implausible in other branches of the consumer literature, in Lancaster's model it does play a central role. Further, it is often true that when multiple varieties are available, the consumer buys amounts of just one at a time. Although this is not excluded by the Lancaster model (this being the case of an optimum at a vertex), the model predicts that this is not a frequent occurrence unless the indifference curves are linear.

Section VI introduces an alternative model of consumer behavior which overcomes this criticism and provides an alternative framework for the analysis of quality change. Basically, this involves the imposition of strong separability-type restrictions on the utility function. These are derived from generalization to a group of goods of the Fisher-Shell "simple repackaging hypothesis." A homothetic category utility function is assumed to exist for a group of goods, defined as the weighted sum of the x, where the weights are interpreted as quality indices. This implies straight line indifference curves between the x. The quality indices are defined to be an a priori unrestricted function of the

characteristics. If a section of the market can be identified where consumers have similar tastes, then the relative values of the quality indices for the purchased goods must be approximately equal to their relative prices. Market competition enforces this. Thus, there is a multiplicative relationship of the form $p_{it}=p_t \cdot g(\text{characteristics})$, in which p_t can be interpreted as a general price index. This results in a pooled time-series/cross-section regression model of $\log p_{it}$ on characteristics and time dummies whose coefficients are interpreted as price indices. It is important to note that the dependent variable *must* appear in *log* form if the regression is to be linear in the parameters. The parameters of the function $g(\ )$ are taste parameters and hence presumed stable. This then corresponds to the second of the two versions of the hedonic technique mentioned earlier.

Reflection will show that the models so far discussed are just relatively refined ways of "judging quality by price." It is paradoxical that the inferences concerning the evaluation of quality which can be drawn from the estimation techniques implied by these models are likely to break down if too many consumers are lazy and *do* judge quality by price. A rational individual would not put his trust in the market mechanism in this way if his marginal rates of substitution (*MRS*) are different from those of others. However, if his *MRS* are different, there are grave problems of aggregation. This is not just a matter of taste differences; I would regard income differences as even more important. The *MRS* which are being implicitly measured depend, in general, on the utility level as well as on tastes. This can be formally introduced into the simple repackaging model by making the function $g(\ )$ depend on the utility level or on past consumption as well as on characteristics. This, however, causes severe problems for the kind of empirical work so far discussed, because the market then no longer forces relative prices to correspond to a unique consumer evaluation of relative characteristics.

The major implication of all this for hedonic studies is that, *at the very least*, careful attention should be paid to cross-sectional disaggregation. As far as possible, markets should be broken into segments based on commodity groupings which make it likely that their consumers have similar *MRS* and these segments should be studied separately. This means that in any sample there will be less cross-sectional variation and hence more problems of multicollinearity. This suggests that investigators will not be able to afford the luxury of intertemporal disaggregation, i.e., separate cross-section regressions for each year. Since this procedure corresponds to the household production model, it appears that the simple repackaging model (which implies pooling time-series and cross-sectional data) represents the more practical theoretical framework.

For some goods a broader research strategy may need to be pursued. In my 1974a paper, I show that in the United Kingdom at least, the cost of living indices of different income groups have increased at different rates (most for the poorest in recent years). This could well be a general problem in some markets for durables. To tackle the problem, it is necessary to combine consumer survey data on durables purchases and ownership with data on prices and characteristics. At least the relevant market segments may then be determined.

In Sections II to VI, it is assumed that consumer preferences are directly reflected in relative prices thus allowing welfare statements about consumers to be deduced. In the tradition of Houthakker, some authors (for example, Phoebus Dhrymes and Makoto Ohta) have taken an opposing view and argued that hedonic

relationships are estimates of manufacturers' pricing strategies or of cost functions. However, Houthakker was not as explicit. He assumes that the consumer is faced with what I would call a "tariff schedule." In the example which he analyzes, the consumer is faced with a base tariff per unit of the good and a second tariff per unit of the characteristic embodied in it. This has been extended by Rosen to a more general shadow price schedule which, he argues, is an equilibrium relationship sustainable in a competitive, well-informed, decentralized market. I have argued in my 1973 paper that the well-known paper by Irma Adelman and Griliches is best interpreted in this general vein. In the present paper, I correct a technical shortcoming in Houthakker and Rosen. This is the assumption that only one variety is bought at a time by one consumer. Only the simple repackaging model yields this behavior as an *outcome* of the optimizing process.

The important point to notice is that if hedonic relationships in part reflect manufacturers' pricing strategies, there is a problem in identifying consumer taste parameters. Since linear indifference loci in x imply corner solutions, the market shadow prices may not reflect those of consumers. These identification problems are not dealt with in this paper. However, they are very important. If the purpose of the hedonic approach is to make welfare comparisons for consumers over time, it is defeated if their marginal rates of substitution cannot be identified.

There is one context in which it may be plausible, however, to argue that consumer tastes dominate. This is in second-hand durables markets where aggregate supply is fairly stable (i.e., approximately exogenous) and particular supplies are usually held in a decentralized fashion. There the models of consumer behavior discussed in this paper are highly relevant if homogeneous market segments can be identified.

The quality problem is a fundamental one in economics. It raises in sharp form the difficulties involved in drawing welfare conclusions from market data. These are problems of interpersonal comparability, of aggregation (see my 1974b paper for a further analysis)—especially of income and taste differences—and of whether market prices represent valid information about rates of substitution and rates of transformation. These are problems which have to be faced in the entire area of aggregative index numbers. They are both interesting and important.

II. The Utility Function and the Household Production Function

I examine a household which purchases market goods in quantities $x_1, \ldots, x_m$ (which yield no direct utility) whose purpose is to jointly produce the commodities Z_1 and Z_2 which yield utility according to[1]

$$U = U(Z_1, Z_2) \tag{1}$$

where $U(\ )$ is assumed to be convex. Let the joint production function be $F(x_1, \ldots, x_m; Z_1, Z_2)=0$. We assume that $F(\ )$ is "neoclassical," i.e., given $x_1, \ldots, x_m$, the production possibility frontier in Z_1 and Z_2 is concave, and the isoquants in $x_1, \ldots, x_m$ given Z_1 and Z_2 are convex. In addition, we assume that the household faces the budget constraint

$$y = \sum p_i x_i \qquad i = 1, \ldots, m \tag{2}$$

The solution to the utility maximization problem can be thought of in two stages.

Stage 1: Minimize $C = \sum p_i x_i$ of producing any given bundle Z_1, Z_2. Let the

[1] The assumption that only two goods are produced and that no other goods enter the utility function is made purely for the sake of simplicity of exposition. The general case when other goods enter $U(\ )$ is briefly discussed below.

Lagrangian be (where x is the vector $(x_1, \ldots, x_m)$)

$$(3) \quad \mathfrak{L}_1 = \sum p_i x_i + \pi(F(x; Z_1, Z_2))$$

Let the solution be given by the cost function $C = C(p; Z_1, Z_2)$ (see R. W. Shepard and Hirofumi Uzawa on the duality between $F(\ )$ and $C(\ )$) where p is the vector $(p_1, \ldots, p_m)$. Define the shadow marginal cost of producing Z_j

$$(4) \quad \pi_j \equiv \frac{\partial C}{\partial Z_j} \quad j = 1, 2$$

Differentiating (3) with respect to Z_j gives

$$(5) \quad \pi_j = \pi \frac{\partial F}{\partial Z_j} \quad j = 1, 2$$

Stage 2: Maximize $U = U(Z_1, Z_2)$ subject to the constraint

$$(6) \quad y = C(p; Z_1, Z_2)$$

Let the Lagrangian problem be

$$(7) \quad \max \mathfrak{L}_2 = U(Z_1, Z_2) + \lambda(y - C(p; Z_1, Z_2))$$

This is illustrated in Figure 1. The frontier AB (at time 0) is obtained by solving $y_0 = C(p_0; Z_1, Z_2)$ for Z_1 in terms of Z_2. Thus AB is concave when $F(\ )$ is neoclassical. The utility maximizing point on AB is Z^0.

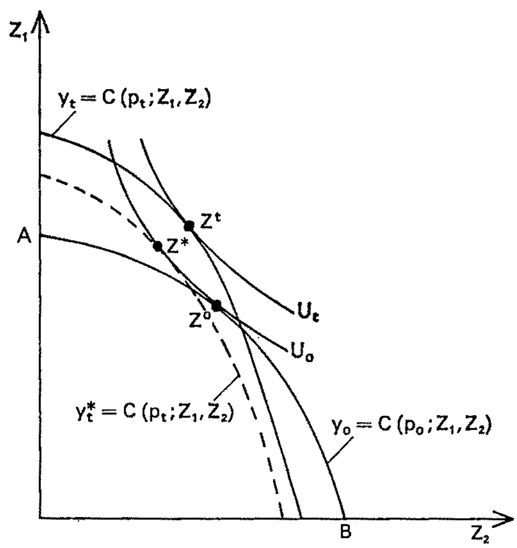


FIGURE 1

The constant utility price index is easily defined in the absence of quality change. Then the form of $C(\ )$ and of $F(\ )$ is unchanged. If expenditure is y_t and prices p_t, the new frontier is given by $y_t = C(p_t; Z_1, Z_2)$. The optimal point on it is Z^t. Define y_t^* as the minimum expenditure to attain U_0 at prices p_t. Then the constant utility price index (with U_0 constant) is given by y_t^*/y_0. This can be explicitly written in terms of the expenditure function $y = \overline{C}(p; U)$ whose functional form depends on those both of the cost function $C(\ )$ and the utility function $U(\ )$. Thus

$$(8) \quad \frac{y^*}{y_0} = \frac{\overline{C}(p_t; U_0)}{\overline{C}(p_0; U_0)}$$

III. Quality Change

When the quality of one or more of the $\{x_i\}$ changes, the household technology changes. The only formal alterations in (8) are the time subscripts t and 0 that must now be given $\overline{C}(\ )$. It is even less plausible that the parameters of $\overline{C}_t(\ )$ could be discovered.

My brief discussion above of the first version of the hedonic technique as used in practice may have suggested that knowledge of $\overline{C}_t(\ )$ is not necessary. If shadow prices π_{1t}, π_{2t} can be estimated, $\overline{C}_t(\ )$ is known to a local linear approximation. Then one might argue that $(\pi_{1t}Z_{10} + \pi_{2t}Z_{20})/y_0$ would at least provide an approximation with a known relationship to the true index. However, the well-known result of the conventional theory of market goods that a Laspeyres index is an upper bound on the true index does not in general hold here. The reasons are twofold. First, it may not be true that $\pi_1 Z_1 + \pi_2 Z_2$ has a stable relationship with market costs. Since

our index is given in ratio form, what we need is for the consistency condition

$$\frac{y_t}{y_s} = \frac{\pi_{1t}Z_{1t} + \pi_{2t}Z_{2t}}{\pi_{1s}Z_{1s} + \pi_{2s}Z_{2s}} \tag{9}$$

to be fulfilled. This condition is stated for a given technology where s and t refer to different time periods. Define a Laspeyres index:

$$\frac{\pi_{1t}Z_{10} + \pi_{2t}Z_{20}}{\pi_{10}Z_{10} + \pi_{20}Z_{20}} \tag{10}$$

The consistency requirement is clearly a minimum requirement for (10) to make much sense.

Now I shall prove the following:

THEOREM: *The consistency condition for the price index is met if $F(x; Z_1, Z_2)$ is homogeneous in both inputs and outputs.*

PROOF:

We have to show that $\sum_{i=1}^{m} p_i x_i = k\sum_j \pi_j Z_j$ where k is independent of Z_j (and hence of p, y), if, and only if, $F(\ )$ is homogeneous in both inputs and outputs.

Giora Hanoch has shown that $F(\ )$ is homogeneous of degree a in inputs and outputs iff

$$\frac{\sum_i \frac{\partial F}{\partial x_i} x_i}{\sum_j \frac{\partial F}{\partial Z_j} Z_j} = a \tag{11}$$

But irrespective of homogeneity, from equation (3) it follows that

$$\frac{\partial F}{\partial Z_j} = \frac{1}{\pi}\frac{\partial C}{\partial Z_j} = \frac{\pi_j}{\pi} \tag{12}$$

$$\frac{\partial F}{\partial x_i} = \frac{p_i}{\pi} \tag{13}$$

Thus, substituting (12) and (13) into (11), we see that homogeneity is equivalent to $\sum_i p_i x_i = a \sum_j \pi_j Z_j$ where a is independent of Z_j, p, and y, which is the desired re-sult. Note that $k=a$.

The theorem requires both homogeneity and that the degree of homogeneity not be affected by quality change. If these conditions are fulfilled, then the constant utility index would be exactly known if the shadow prices at Z^0 and Z^* were known.

The second reason why the relationship of (10) to y_t^*/y_0 is in general unknown is that, even if observable, the shadow prices π_{1t}, π_{2t} correspond to the *actually* bought bundle Z^t rather than the *hypothetical* bundle Z^*. Figure 2 suggests that (10) could lie above *or* below $y_t^* y_0$. The tangent at Z^t and the dotted lines through Z^0 and Z^{00} have slope $-\pi_{2t}/\pi_{1t}$. The actual shadow costs at Z^* may be such that $\pi_{1t}Z_1^*+\pi_{2t}Z_2^*$ could be greater than or less than y^*. Even if $y^*=\pi_{1t}Z_1^*+\pi_{2t}Z_2^*$, then if Z^0 were the period 0 constraint, (10) exceeds y_t^*/y_a, while if Z^{00} were the period 0 constraint, (10) lies below y_t^*/y_a.

The next section shows that there are two sets of circumstances under which this problem is resolved. Also to be faced are three further practical difficulties: the

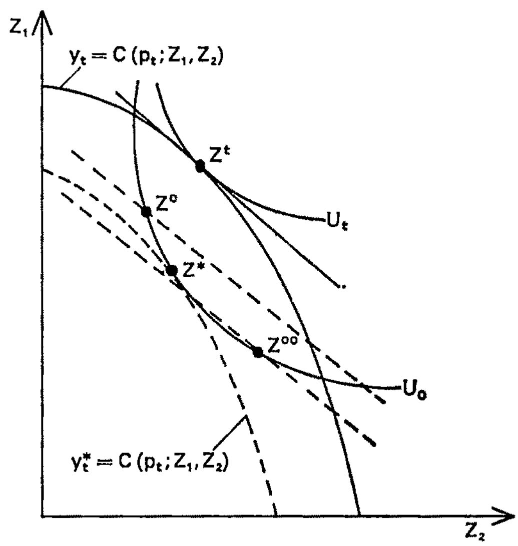


FIGURE 2

measurement of Z_1 and Z_2, the observation of the shadow costs π_1 and π_2, and aggregation across consumers.

IV. When Do π_{1t}, π_{2t} Correspond to Current Prices and Base Utility?

I shall show that there are two alternative conditions which guarantee this. As we have seen, the problem is that the shadow costs appropriate to Z^* may not be those actually observable at Z^t. This problem would not arise if *either* (a) relative shadow costs are independent of Z *or* (b) given constant returns, the optimum points at given prices p and given technology, all lie on one expansion path. The discussion above relating to Figure 2 suggests that one requirement for the resolution of the above problem is that the transformation locus be a straight line.

It is possible to show the following:

THEOREM: *A necessary and sufficient condition on the technology for* π_1/π_2 *to be independent of* Z_1 *and* Z_2 *(and of* U *and* C*) is that the cost function has the form* $C = C(\alpha_1 Z_1 + \alpha_2 Z_2, p)$ *where* α_j *depends only on* $p_1, \ldots, p_m$.

PROOF:

Sufficiency is obvious; we concentrate on necessity. The condition,

$$\text{(14)} \quad \frac{\partial C/\partial Z_1}{\partial C/\partial Z_2} = \frac{\pi_1}{\pi_2} = g(p_1, \ldots, p_m)$$

implies a straight-line transformation locus. For a fixed value $C = C'$, the equation of this locus is

$$\text{(15)} \quad C' = \bar{\alpha}_1 Z_1 + \bar{\alpha}_2 Z_2$$

$$\text{where} \quad \frac{\bar{\alpha}_1}{\bar{\alpha}_2} = \frac{\pi_1}{\pi_2}$$

Since the same is true of any monotonic function, $C(C')$ where $G(\ )$ can depend on p, the general solution to equation (14) and $C = C(p; Z_1, Z_2)$ is

$$\text{(16)} \quad C = G(\alpha_1 Z_1 + \alpha_2 Z_2, p)$$

If $F(\ )$ is homogeneous of degree a, $C(p; Z_1, Z_2)$ is homogeneous of degree $1/a$. Then (16) takes the form

$$\text{(17)} \quad C = h(\alpha_1 Z_1 + \alpha_2 Z_2)^{1/a}$$

$$\text{and} \quad \frac{\alpha_1}{\alpha_2} = \frac{\pi_1}{\pi_2} \quad \text{and} \quad h = h(p)$$

By homogeneity, Euler's equation implies

$$h(\alpha_1 Z_1 + \alpha_2 Z_2)^{1/a} \equiv a(\pi_1 Z_1 + \pi_2 Z_2)$$

Thus

$$\text{(18)} \quad \pi_j = \frac{h}{a} \alpha_j (\alpha_1 Z_1 + \alpha_2 Z_2)^{1/a - 1}, \qquad j = 1, 2$$

At first sight[2] it appears as if in the absence of restrictions on tastes, homogeneity together with a straight line transformation locus, i.e., a cost function of the form (17), ought to be enough to guarantee that the Laspeyres index (10) is an upper bound on the true index (8). It is certainly clear that irrespective of technology,

$$\text{(19)} \quad \frac{C_t(p_t; Z_{10}, Z_{20})}{C_0} \geq \frac{\overline{C}_t(p_t; U_0)}{C_0}$$

since (Z_{10}, Z_{20}) yields utility U_0, though not necessarily in the cheapest way. However, given the technology of (17), the following *reverse* inequality may be true:

$$\text{(20)} \quad \frac{\pi_{1t} Z_{10} + \pi_{2t} Z_{20}}{\pi_{10} Z_{10} + \pi_{20} Z_{20}} < \frac{h_t(\alpha_{1t} Z_{10} + \alpha_{2t} Z_{20})^{1/a}}{C_0} = \frac{C_t(p_t; Z_{10}, Z_{20})}{C_0}$$

Substituting from (18) and using $C_0 = h_0 \cdot (\alpha_{10} Z_{10} + \alpha_{20} Z_{20})^{1/a}$, it is easy to show that the inequality in (20) is satisfied with increasing returns to scale (i.e., $a > 1$) if $Z_{1t} > Z_{10}$ and $Z_{2t} > Z_{20}$, and with diminishing

[2] I am very grateful to Pollak for saving me from an error on this point.

returns to scale if $Z_{1t} < Z_{10}$ and $Z_{2t} < Z_{20}$.

The economic interpretation of this is that the shadow costs fall with output expansion under increasing returns and output contraction under diminishing returns. Therefore it is possible for the Laspeyres index (10) to lie *below* the true index (8). It is clear therefore that the further restriction of homogeneity of degree one, i.e., constant returns, must hold if this problem is to be resolved. Then the cost function can be written in the form, say,

$$C = h\alpha_1 Z_1 + h\alpha_2 Z_2 = \hat{C}_1(p)Z_1 + \hat{C}_2(p)Z_2 \tag{21}$$

Robert Hall (1973) drawing on Samuelson (1966) has shown that (21) is equivalent to constant returns and nonjointness. Nonjointness in the inputs $x_1, \ldots, x_m$ means that the joint production function $F(x_1, \ldots, x_m; Z_1, Z_2) = 0$ can be written as

$$\begin{aligned} Z_1 &= f_1(x_{11}, \ldots, x_{m1}) \\ Z_2 &= f_2(x_{12}, \ldots, x_{m2}) \end{aligned} \tag{22}$$

where $x_i = x_{i1} + x_{i2}$ for $i = 1, \ldots, m$

Under constant returns and nonjointness, the shadow costs are independent of Z_1, Z_2.[3] This is essentially the nonsubstitution theorem of Paul Samuelson (1951). In the current context, apart from providing a sufficient condition for the Laspeyres index of shadow costs to be an upper bound on the true index, it has the additional advantage of solving the problem of aggregation across consumers: as long as each faces the same technology, each will face the same shadow costs irrespective of tastes. The interest of form (16) is that for *relative* shadow costs, the aggregation problem is solved without nonjointness having to be assumed. However, neither (16) nor (21) are at all general. Both, for example, are inconsistent with the characteristics approach of Lancaster (see below) in which jointness is at the core.

Although the gain from being able to assume nonjointness is thus considerable, it is often implausible for household production and rather demanding in terms of information on internal allocation.[4] The problem for hedonic work is not only in measuring the shadow costs but in identifying Z. It is implausible to regard Z as being embodied in the goods. Nonjointness implies that the level of each input can essentially be split into separate parts, each of which is used in only one branch of production. Thus if $x_i = x_{i1} + x_{i2}$, the production of Z_1 depends on x_{i1} and that of Z_2 on x_{i2}. Many types of consumption are not, however, of this type. For example, the use of an automobile provides travel services and psychic rewards jointly.

Fortunately, there is a second, less restrictive condition which although it does not solve the aggregation problem is sufficient for the Laspeyres index always to be an upper bound on the constant utility index. Given constant returns of the household production function, this is simply that the utility function be homothetic.

[3] The same conclusions follow if instead of market good x_1 we have labor in the following model:

$$\begin{aligned} Z_1 &= f_1(L_1, x_{21}, \ldots, x_{m1}) \\ Z_2 &= f_2(L_2, x_{22}, \ldots, x_{m2}) \end{aligned}$$

and the budget constraint is replaced by

$$\sum_{i=2}^{m} p_i x_i = w(\bar{L} - L_1 - L_2)$$

where $\bar{L}$ is the total time which can be allocated either to household production or paid employment and where w is the wage rate.

[4] A point about allocative Pareto efficiency can also be made: although household production is a nonmarket activity, given nonjointness and constant returns, the same shadow prices rule in different households. This is a result rather like the factor price equalization theorem. Much of the alleged usefulness in terms of potential empirical predictions of the household production approach rests on this result. Drawing on Hall's result, see (21), Pollak and Michael Wachter point out this approach has little content if the shadow prices vary across households. Thus much of the literature based on this approach rests on very shaky empirical foundations.

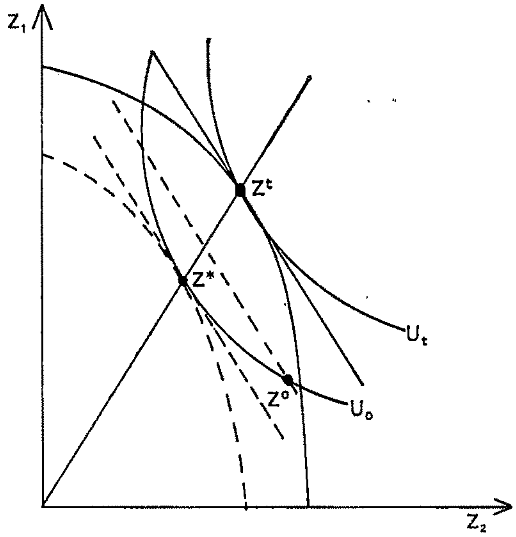


FIGURE 3

This guarantees that given prices and given technology, the relative shadow costs are the same whatever the level of consumption. The reason is obvious: constant returns in production and homothetic preferences together ensure that the optimal points will always be on the same expansion paths and hence correspond to the same marginal rate of transformation. A glance at Figure 3 will convince the reader that then the problem revealed in Figure 2 disappears: if Z^0 can be bought at π_{1t}, π_{2t} then Z^* can *always* also be bought. Note that again we need constant returns rather than merely homogeneity so that the total shadow cost just equals the market cost. Homotheticity, i.e., unitary income elasticities of demand for the Z is also a highly restrictive requirement which is usually refuted by empirical work.

Introducing other market goods does not create serious problems. There are two cases: separability and nonseparability.

Separability: We write the utility function

$$U = U(X_0(Z_1, Z_2), X_1, \ldots, X_n)$$

Then the consumer chooses in two stages: first how to allocate his budget among X_0, $X_1, \ldots, X_n$ and given the budget for X_0, how to choose the optimal values of Z_1 and Z_2. Then a category price index can be defined as the expenditure necessary to reach a given level of category utility X_0. Such a category price index is easily incorporated into an overall price index.

Nonseparability: a proper subindex cannot in general be defined when[5]

$$U = U(Z_1, Z_2, X_1, \ldots, X_n) \tag{23}$$

However, with nonjointness and constant returns, an upper bound as before is given by:

$$\frac{a \sum_{j=1}^{2} \pi_{jt} Z_{j0} + \sum_{i=1}^{n} P_{it} X_{i0}}{a \sum_{j=1}^{2} \pi_{j0} Z_{j0} + \sum_{i=1}^{n} P_{i0} X_{i0}} \tag{24}$$

Increasing the number of Z beyond two does not affect any of the substantive results of this paper.

V. Some Special Cases

The best known model of household production is that of Lancaster. Here we have[6]

$$\begin{aligned} Z_1 &= \sum b_{1i} x_i \\ Z_2 &= \sum b_{2i} x_i \end{aligned} \tag{25}$$

Each unit of good i is made up of a fixed amount b_{1i} of Z_1 and b_{2i} of Z_2.

The transformation locus corresponding to a given income and market prices can be represented as in Figure 4.

For any given bundle Z_1, Z_2, it can be assumed that the consumer will have minimized the cost of purchasing that bundle subject to $x_i \geq 0$ at the given prices p_i.

[5] This is a well-known point, see Gorman (1961). For a particularly lucid explanation, see Pollak.

[6] We do not make Lancaster's distinction, here inessential, between market goods and "activities." Note by the way that the technology reveals constant returns.

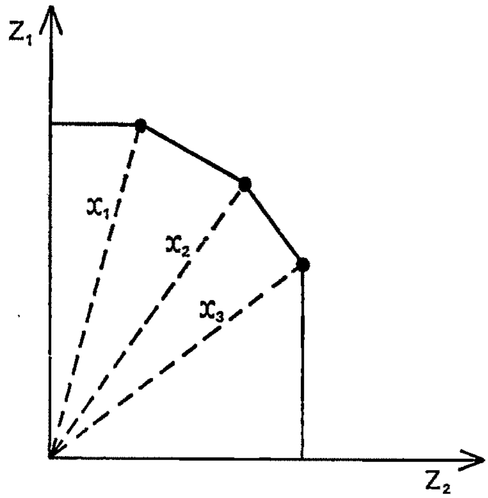


FIGURE 4

Given divisibility, shadow prices π_1 and π_2 corresponding to this linear programming problem can be calculated. As Richard Lipsey and Gideon Rosenbluth show, for purchased goods

$$p_i = \pi_1 b_{1i} + \pi_2 b_{2i} \tag{26}$$

In principle, linear programming methods could be used to compute the π_j. However, a serious difficulty is likely to be that the solution will be sensitive to the exact way the problem is specified: information on the b_j will never be available in enough detail and with enough accuracy. The form of (26) also suggests a regression approach. Because some sort of average relationship, though ideally disaggregated according to market segments, is then estimated and the b_j are typically somewhat collinear, the omitted or misspecified variable problem is likely to be less serious. Thus two problems mentioned above, i.e., measuring Z_j and estimating π_j appear to be solved.[7]

There remains the problem of aggregating across consumers. Clearly, if consumers' indifference curves are similar enough[8] in shape, they will all choose points corresponding to the same π_1/π_2. Then, abstracting from information problems, for the sellers of a good to sell anything, they must price according to the relationship $p_i = \pi_i b_{1i} + \pi_2 b_{2i}$. In a way, the identity of tastes enforces a strong form of perfect competition.

If the indifference curves are not similar enough, different people will choose points corresponding to different values of π_1/π_2. In this case if one nevertheless wants to pick up some sort of median relationship, there is an argument for weighting the observations by value shares. However, for the poorer members of the community, the price quality relationships may well be substantially different and it is difficult to know what meaning to attach to the estimated relationships.

Suppose now that the production function is additive but not linear. For example, suppose[9]

$$\begin{aligned} Z_1 &= \sum b_{1i} \log x_i \\ Z_2 &= \sum b_{2i} \log x_i \end{aligned} \tag{27}$$

Here

$$x_i p_i = \pi_1 b_{1i} + \pi_2 b_{2i} \tag{28}$$

It is no longer so plausible to regard b_{ji} as the amount of the jth characteristic

[7] Incidentally, it may be that one unit of the good, irrespective of how much b_1 and b_2 are embodied in it, has an intrinsic value. In that case, $p_i = \pi_0 + \pi_1 b_{1i} + \pi_2 b_{2i}$, which is a form often fitted. Here the utility function is $U = U(Z_0, Z_1, Z_2)$ where Z_0 is the number of units of market goods consumed.

[8] This not only implies similarity of tastes; if tastes are distinctly nonhomothetic, then incomes would have to be similar as well.

[9] Gorman in a regrettably unpublished paper from 1956 anticipates Lancaster's work to a remarkable degree. Gorman considered the two cases, i.e., $Z_j = \sum b_{ji} x_i$ and $Z_j = \sum b_{ji} \log x_i$. However, he assumed that b_{ji} are unobservable. He proposed, in great detail, that factor analysis could be used to estimate simultaneously (at least up to constants of proportionality) b_{ji} and π_j. In a sense, he can be said to have solved a problem more difficult still than that solved by the hedonic technique, well before the latter achieved any currency. His suggestion of factor analysis could also be useful in analyzing the residuals for regressions where it is thought that not all the relevant b_{ji} were included.

embodied in the ith market good. If b_{ji} is the horsepower of the model, it is now no longer total horsepower which is relevant, but the sum of horsepower weighted by the *logs* of the quantities purchased. Notice that in this form, the relationships between the prices of the market goods and b_{ji} involve the quantities purchased, x_i.[10] Not only does this cause severe aggregation problems, but since the technology is not constant returns, the Laspeyres upper bound may not hold.

The most popular empirical form for hedonic regressions has been of $\log p_i$ on $\sum \pi_j b_{ji}$. It is easy to see that the following impossibility theorem holds:

THEOREM: *The semilog form of the price/characteristics relationship is not possible in the household production framework.*

PROOF:

The most general way of formulating models like (25) and (27) to incorporate information from the characteristics is as $Z_j = f_j(b_{j1}, \ldots, b_{jm}; x_1, \ldots, x_m)$.

$$p_i = \pi_1 \frac{\partial f_1}{\partial x_i} + \pi_2 \frac{\partial f_2}{\partial x_i}$$

Thus we need

$$\pi_1 \frac{\partial f_1}{\partial x_i} + \pi_2 \frac{\partial f_2}{\partial x_i} \equiv e^{\pi_1 b_{1i} + \pi_2 b_{2i}} \tag{29}$$

This cannot be satisfied for any nontrivial version of the function $f_j(\ )$ because an additive form in π_1, π_2 cannot be identical to a multiplicative form in π_1, π_2, i.e., $a\pi_1 + b\pi_2 \neq A(\pi_1) \cdot B(\pi_2)$.

The nonjoint model is no more yielding in this respect. Also, the nonjoint model of household production suffers from the handicap that the allocation of x_i within the household and hence Z_j are typically unobservable.[11]

VI. An Alternative Approach: The Simple Repackaging Hypothesis

Outside the household production framework, and in particular the Lancaster model, there is only *one* other simple theoretical model of consumer behavior in which the hedonic technique as currently practiced can be interpreted as a direct reflection of consumer behavior.

Under this hypothesis, it is assumed that each market good has a quality index which is a function of a set of physical characteristics. This relationship is the same for all market goods of a general type (say, refrigerators or some grouping of refrigerators) and being a question of tastes, is independent of market variables. Under this assumption, which is essentially the Fisher-Shell assumption of "simple repackaging" quality differences and quality change, *market* goods of a given type can be aggregated; the aggregate is simply the sum of the quality indices weighted by the number of units of each good purchased. Formally, the utility function can be written[12]

$$U = U\left\{X_0\left(\sum_{i=1}^{m} a_i x_i\right), X_1, \ldots, X_n\right\} \tag{30}$$

where $X_0(\ )$ is the category utility function for the group $(x_1, \ldots, x_m)$ and X_s, $s = 1, \ldots, n$, are other market goods or category functions for groups of other market goods.

The a_i are quality indices and may be made a function of the characteristics (b_{1i}, b_{2i}),

[10] See my 1972 paper for a similar result in a different theoretical framework.

[11] However, as we shall see in Section VII, the semilog model can be interpreted in a different way. It must also be pointed out that a *partial* relationship of a semilog form between the price and one characteristic is quite permissible. This is because the units in which b_{ji} are measured are arbitrary. It may be that the appropriate measure is $b'_{ji} = \exp b_{ji}$. Then the partial relationship between p_i and the originally measured characteristic b_{ji} is indeed semilog. Indeed the Lancaster model is fairly general in this way; characteristics can, for example, be defined in terms of the interactions of the actually measured characteristics.

[12] A similar formulation to analyze quality change is proposed in a working paper by Hugh Davies.

$$a_i = g(b_{1i}, b_{2i}) \tag{31}$$

For empirical work, a particular form for $g(\ )$ has to be put forward whose choice is mainly an empirical question. If $g(\ )$ is exponential with a constant term, then the popular semilog regression form results. However, since the a_i are parameters of the utility function, they must be permitted to be stable over the medium term at least, or else made stable functions of utility or lagged consumption or income. This is quite different from the usually quite abrupt changes from year to year which are implied by the single year regression method recommended by Griliches (1971).

In a sense, the highly restrictive implication of linear additivity of the category utility function has an attractive property. This is that if for one good k, $a_k/p_k > a_i/p_i$ for $i=1, \ldots, m$, $i \neq k$, only good k is bought. If $a_k/p_k = a_i/p_i$ for one or more $i \neq k$, the distribution of demand between these goods and good k is indeterminate. It seems to be quite frequently observed that only one variety of a good is bought by one consumer at any one time. Thus rather plausible discontinuities in individual behavior when the prices p_i or quality indices a_i change, or when new goods appear, are a consequence of this model.[13]

If everyone has similar tastes, then it is assumed that competition will force relative prices to be approximately equal to relative quality indices at each point in time, i.e., p_i will be approximately proportional to $g(b_{1i}, b_{2i})$. Over time, the general level of prices changes and the constant of proportionality can be interpreted as a price index. Thus

$$p_{it} = \bar{p}_t g(b_{1it}, b_{2it}) \tag{32}$$

This becomes a linear regression model after a *log* transformation. It is estimated on pooled time-series/cross-section data and corresponds to the second variant of the hedonic technique mentioned in the introduction. Thus $log\ p_{it} = log\ \bar{p}_t + log\ g(b_{1it}, b_{2it})$. The $log\ \bar{p}_t$ can be picked up as the coefficients of time dummies. This model can be generalized to used durables when in logarithmic terms there is also an additive age effect. This was done by Hall (1969) in formalizing Philip Cagan's earlier work.

Of course, it is not plausible that everyone has the same tastes. Families have different compositions and it is unlikely that they would have similar preferences, for example with respect to roominess versus acceleration in automobiles. Perhaps even more serious is the plausible situation that a_i or (which is equivalent) the function $g(\ )$ are nonhomothetic, i.e., depend on utility U or at least category utility X_0.[14] This does not invalidate the linearity in x of the indifference loci or the discontinuous behavior which was discussed above. But it implies that for different income levels consumers will have different marginal rates of substitution. It therefore calls into question the assumption that competition will force relative prices into equality with relative quality indices. If quality means different things at different levels of income, we have a serious problem of how to interpret the above regression model based on aggregated market data. That this is in some respects part of the general problem of what aggregate index numbers are supposed to represent (see my 1974b paper) is no ground for complacency. At the very least, some cross-sectional disaggregation ought to be tried in order to attempt to

[13] Note that the parameters in the utility function corresponding to new goods are known given $g(\ )$ as soon as the specifications are known. A similar property but with non-linear indifference curves results from my extension of "simple" to "variable repackaging" in my 1972 paper. However, predicted behavior is then continuous.

[14] Davies, in a revised version of his paper, has formulated such a nonhomothetic model with some particular functional forms.

separate out different market segments. However, this may not be enough and budget data may have to be analyzed. An attempt to do this, though in the context of the Lancaster model rather than the present one, has been made by William Alexander.

It is clear that *in general*, if the Lancaster model applies, quality change is not of the simple repackaging type. Suppose the market goods are refrigerators and Z_1 is size and Z_2 minimum temperature. It is quite likely that quality change, say, a lower minimum temperature, will augment the value of some other good, say ice cream. This is not equivalent to "more" of a refrigerator, though for a consumer with a given income at a given price configuration there will no doubt be a determinate tradeoff between more and having a lower minimum temperature. The point is that as soon as income and prices change, this tradeoff will be different. However, the gain in generality of the Lancaster model relative to simple repackaging is not great. This is because if there are other market goods, we still need separability for the Laspeyres upper bound to hold.

Only if the utility function is separable and the category function $X_0(\ )$ is linear and additive both in the x_i, $i=1, \ldots, m$ and the Z_j, $j=1, 2$, do the Lancaster model and the simple repackaging model coincide. Then

$$\sum a_i x_i \equiv \gamma_1 Z_1 + \gamma_2 Z_2 \tag{33}$$

where γ_1, γ_2 are constants and

$$a_i = \gamma_1 b_{1i} + \gamma_2 b_{2i} \tag{34}$$

Thus the hedonic regressions have the form

$$p_{it} = \bar{p}_t(\gamma_1 b_{1it} + \gamma_2 b_{2it}) \tag{35}$$

or

$$p_{it} = \bar{p}_t(\gamma_0 + \gamma_1 b_{1it} + \gamma_2 b_{2it}) \tag{36}$$

if as raised in footnote 7 a unit of x_i, $i=1, \ldots, m$, has a value independent of how much it possesses in characteristics. The γ_j are constant over time. Incidentally (35) and (36) raise some estimation problems since an *additive* time dummy cannot pick up the quality corrected price index $\bar{p}$.[15]

Since this section has been about various sorts of separability, this seems a good place to comment on Muth. His paper is less about household production than about separability. Muth assumes

$$Z_1 = f_1(x_1, \ldots, x_s)$$
$$Z_2 = f_2(x_{s+1}, \ldots, x_m)$$

where $f_1(\ )$ and $f_2(\ )$ reveal constant returns. Clearly this is the well-known case of homothetic separability and Muth's substantive results are a development of results by Robert Strotz, Michio Morishima, and Gorman (1961). A similar point is made by Patrick Geary and Morishima.

VII. The Houthakker Approach

Having reached this point, it is now possible to discuss the classic paper by Houthakker.[16] Adelman and Griliches claim this paper as an antecedent of theirs. However, as Robert Lucas[17] has pointed out, their attempt to base the hedonic ap-

[15] The technique of picking up price indices as the coefficients of time dummies in *additively linear* time-series cross-section analysis has a different theoretical justification. It corresponds to the Lancaster model where $Z_0=\sum x_i$, $Z_j=\sum b_{ji}x_i$, $j=1, \ldots, k$, and where the relative values of $\pi_1, \ldots, \pi_k$ are constant over time and only π_0 changes. This implies a utility function of the form

$$U = U\left(Z_0, \sum_{j=1}^{k} \gamma_j Z_j, X_1, \ldots, X_n\right)$$

where possibly the first two elements in $U(\ )$ are further separable from $X_1, \ldots, X_n$. However, the corresponding index numbers have not usually been correctly calculated.

[16] It is a classic in part because it has one of the earliest expositions of the duality approach in consumer theory.

[17] I thank Griliches for bringing this reference to my attention.

proach in consumer theory has some serious gaps—another reason for going back to Houthakker.

Reducing his model to essentials and giving a concrete illustration we have the utility function

$$U = U(x, b, X_1, X_2, \ldots, X_n) \tag{37}$$

where x refers to the quantity of (say) auto services consumed with (say) horsepower b. Suppose that the other goods $X_1, \ldots, X_n$ are available in only one variety each. Assume that all auto services actually consumed have the same horsepower, i.e., that only *one* variety is bought, but that a continuum of services with different amounts of horsepower are available.

The crucial assumption which is not justified further, is that *the consumer faces a tariff schedule quite outside his control.* This relates the price of the unit to horsepower and essentially is a two-part tariff of the form

$$p = \pi_0 + \pi_b b \tag{38}$$

Houthakker calls π_0 the "quantity price" and π_b the "quality price."

Thus the maximization problem is subject to *two* constraints: the budget constraint and the tariff schedule.

Maximize $U = U(x, b, X_1, X_2, \ldots, X_n)$ subject to

$$\begin{aligned} y &= px + \sum_{i=1}^{n} P_i X_i \\ p &= \pi_0 + \pi_b b \end{aligned} \tag{39}$$

If we forget about the other goods, $X_1, \ldots, X_n$, this can be represented in Figure 5. Note that for a unique interior solution, the optimal indifference curve must not be "less convex" than the tariff schedule.

The fact, as Houthakker states, that the arguments in the indirect utility function are π_0, π_b, $P_1, \ldots, P_n$, y, can be given

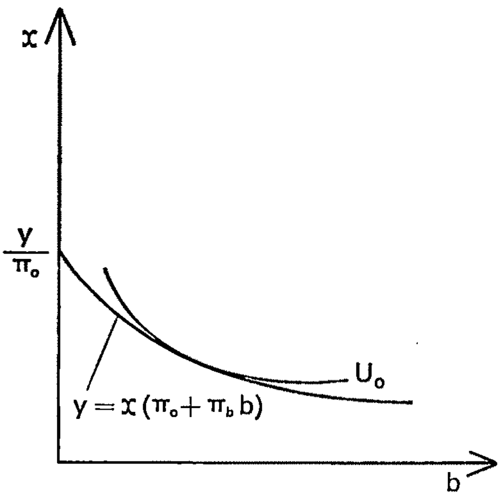


FIGURE 5

intuitive support by the following argument:

Define $bx = Z$ and redefine the utility function to be

$$U = \overline{U}(x, Z, X_1, X_2, \ldots, X_n)$$

Z is just like a conventional good with a price π_b. Thus, the expenditure function through which the cost of living index and its approximations can be defined has the form

$$y = m(\pi_0, \pi_b, P_1, \ldots, P_n, U)$$

This presentation of Houthakker's argument clarifies, I hope, a subtle confusion in much of the hedonic literature. My use of the term "tariff schedule" is related to Richard Stone's discussion of transport prices in a quality context. He argues that often the price can be thought of as consisting of a ton/mile component and a loading or unloading component. Here then, the so-called shadow prices are unambiguously given to the consumer and can be thought of as primarily production cost determined. It is often not clear in later literature whether shadow prices are internal or external to household decision

making. The Adelman and Griliches paper is a case in point (see my 1973 paper).

The extension of Houthakker's analysis to several characteristics is trivial as long as the tariff schedule maintains its linear form. A further extension is contained in the recent paper by Rosen. Rosen *assumes* the existence of a general tariff schedule $p_i = p(b_{1i}, \ldots, b_{ki})$ where b_{ji} is the amount per ith good of the jth characteristic. Given its existence, he tells an analytically rich story of consumers with different tastes and producers with different technologies in a competitive, well-informed market context, which explains how such a tariff schedule may be sustained.

One uncomfortable feature of the papers both of Houthakker and Rosen is the assumption that only one variety is bought. This is imposed as an extraneous *assumption* rather than as the *outcome* of the optimizing model.

Fortunately, however, the discussion of the previous section at once reveals the solution to this problem. The answer is quite simply that the utility function of the individual consumer *must have* the form (30). The reason is that within the category of goods with which we are concerned, if only one variety is bought, we must have a corner solution. Now the only sort of indifference curve which will always produce a corner solution when faced with a linear budget constraint is a linear one (we assume that concave indifference curves are ruled out). The form (30) exactly corresponds to the two-stage decision process which is really implied by Houthakker and Rosen. At one stage the choice of variety is made, at the other the choice of how much is to be spent on that category and hence on that variety.

However, as we saw in Section VI, there is a way of generalizing (30). This is to make the parameters $\{a_i\}$ functions of the level of utility or category utility. Then the marginal rates of substitution differ for different levels of income. For Rosen this would have the analytical advantage of providing a natural motivation for what he calls "taste differences," i.e., differences in the relative values of the $\{a_i\}$ for different consumers. The existence of smooth distributions of such differences plays an important role in his analysis.

Even though I have succeeded in clarifying the consumer theory implied in the papers of Houthakker and Rosen, the point remains that their shadow prices do not directly reflect the $\{a_i\}$. Instead their shadow prices are supply or market generated in some way. If consumers adjust fully, their shadow prices in utility terms should reflect these market shadow prices if the latter exist. A central problem of the exercise then is to which social group, if any, the measured price index corresponds.

REFERENCES

I. **Adelman and Z. Griliches**, "On an Index of Quality Change," *J. Amer. Statist. Assn.*, Sept. 1961, *56*, 535–48.

W. E. **Alexander**, "The Hedonic Theory and Measurement of Quality Change," paper presented at the Canadian Economics Assn. Meetings, June 1973, Kingston, Ontario.

G. S. **Becker**, "A Theory of the Allocation of Time," *Econ. J.*, Sept. 1965, *75*, 493–517.

P. **Cagan**, "Measuring Quality Change and the Purchasing Power of Money: An Exploratory Study of Automobiles," *Nat. Banking Rev.*, Dec. 1965, *3*, 217–36; reprinted in Z. Griliches, ed., *Price Indexes and Quality Change: Studies in New Methods of Measurement*, Cambridge, Mass. 1971.

H. B. **Davies**, "The Consumer's Choice Among Qualities of Goods," Fels disc. pap. 25, Univ. of Pennsylvania, Nov. 1972; disc. pap. no. 26, Birkbeck College, Mar. 1974.

P. J. **Dhrymes**, "Price and Quality Changes in Consumer Capital Goods: An Empirical Study," in Z. Griliches, ed., *Price Indexes and Quality Change: Studies in New*

Methods of Measurement, Cambridge, Mass. 1971.

F. M. Fisher and K. Shell, "Taste and Quality Change in the Pure Theory of the True Cost of Living Index," in Z. Griliches, ed., *Price Indexes and Quality Change: Studies in New Methods of Measurement*, Cambridge, Mass. 1971.

P. T. Geary and M. Morishima, "Demand and Supply under Separability," in M. Morishima et al., eds., *The Theory of Demand: Real and Monetary*, Oxford 1973.

R. J. Gordon, "The Measurement Bias in Price Indexes for Capital Goods," *Rev. Income and Wealth*, June 1971, *17*, 121–75.

W. M. Gorman, "A Possible Procedure for Analysing Quality Differentials in the Egg Market," with addendum, mimeo., 1956.

———, "Separability and Aggregation," *Econometrica*, 1961, *29*, 499–516.

Z. Griliches, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," (1961) in his *Price Indexes and Quality Change: Studies in New Methods of Measurement*, Cambridge, Mass. 1971.

———, "Hedonic Price Indexes Revisited: Some Notes on the State of the Art," (1967) rev. in his *"Price Indexes and Quality Change: Studies in New Methods of Measurement*, Cambridge, Mass. 1971.

———, *"Price Indexes and Quality Change: Studies in New Methods of Measurement*, Cambridge, Mass. 1971.

R. E. Hall, "The Measurement of Quality Change from Vintage Price Data," in Z. Griliches, ed., *Price Indexes and Quality Change: Studies in New Methods of Measurement*, Cambridge, Mass. 1971.

———, "The Specification of Technology with Several Kinds of Output," *J. Polit. Econ.*, July/Aug. 1973, *81*, 878–92.

G. Hanoch, "Homotheticity in Joint Production," *J. Econ. Theory*, Dec. 1970, *2*, 423–26.

H. S. Houthakker, "Compensated Changes in Quantities and Qualities Consumed," *Rev. Econ. Stud.*, 1952, *19*, 155–64.

K. Lancaster, "A New Approach to Consumer Theory," *J. Polit. Econ.*, Apr. 1966, *74*, 132–57.

R. G. Lipsey and G. Rosenbluth, "A Contribution to the New Theory of Demand: A Rehabilitation of the Giffen Good," *Can. J. Econ.*, May 1971, *4*, 131–63.

R. E. B. Lucas, "Working Conditions, Wage-Rates and Human Capital: A Hedonic Study," unpublished doctoral dissertation, M. I. T., Oct. 1972.

D. McFadden, *An Econometric Approach to Production Theory*, Amsterdam, forthcoming.

M. Morishima, "The Problem of Intrinsic Complementarity and Separability of Goods," *Metroeconomica*, Dec. 1959, *11*, 188–202.

——— et al., *The Theory of Demand: Real and Monetary*, Oxford 1973.

J. Muellbauer, book review of *Price Indexes and Quality Change: Studies in New Methods of Measurement*, *J. Econ. Lit.*, Sept. 1972, *10*, 830–32.

———, "Characteristics, Substitution Between Goods and Quality," mimeo., Birkbeck College, Nov. 1972.

———, "The Divisia Index Approach to Quality Change: A Note on Adelman and Griliches," disc. pap. no. 9, Birkbeck College, Oct. 1973.

———, (1974a) "Prices and Inequality: The Recent U.K. Experience," *Econ. J.*, Mar. 1974, *84*, 32–55.

———, (1974b) "The Political Economy of Price Indices," disc. pap. no. 22, Birkbeck College, Mar. 1974.

R. F. Muth, "Household Production and Consumer Demand Functions," *Econometrica*, July 1966, *34*, 699–708.

M. Ohta, "Hedonic Price Index for Boiler and Turbo-Generator: A Cost Function Approach," unpublished doctoral dissertation, Univ. California, Berkeley, 1971.

R. A. Pollak, "Subindexes of the Cost of Living Index," res. disc. pap. 15, Office of Prices and Living Conditions, Bureau of Labor Statistics, June 1973.

——— and M. L. Wachter, "The Relevance of the Household Production Function and Its Implications for the Allocation of Time," *J. Polit. Econ.*, forthcoming.

S. Rosen, "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Com-

petition," *J. Polit. Econ.*, Jan./Feb. 1974, *82*, 34–55.

P. Samuelson, "The Fundamental Singularity Theorem for Non-Joint Production," *Int. Econ. Rev.*, Jan. 1966, *7*, 34–41.

———, "Abstract of a Theorem Concerning Substitutability in Open Leontief Models," in T.C. Koopmans, ed., *Activity Analysis of Production and Allocation*, New York 1951.

R. Shephard, *Theory of Cost and Production Functions*, Princeton 1953.

R. Stone, *Quantity and Price Indexes in National Accounts*, Paris 1956.

R. H. Strotz, "The Empirical Implications of a Utility Tree," *Econometrica*, Apr. 1957, *25*, 269–80.

J. E. Triplett, "The Theory of Hedonic Quality Measurement and Its Use in Price Indexes," Bureau of Labor Statistics staff pap. 6, Washington 1971.

H. Uzawa, "Duality Principles in the Theory of Cost and Production," *Int. Econ. Rev.*, May 1964, *5*, 216–20.

Fairness and Envy

By ALLAN FELDMAN AND ALAN KIRMAN*

"Envy's a stronger spur than pay."
John Gay

"This only grant me that my means may lie Too low for envy, for contempt too high."
Abraham Cowley

Standard neoclassical economic analysis is typically concerned with individual utility maximization. In this paper we shall consider a problem of constrained social welfare maximization. Our criterion of social welfare is "fairness," and we shall discuss how this may be maximized by a move from an initial allocation to a final fairer allocation, subject to the constraint that no one be made worse off by the move. We think the goal of fairness maximization characterizes, albeit in a simplistic way, the goals pursued by "enlightened" governments in their redistributional policies. We shall also discuss a concept of complete fairness and illustrate some of its weaknesses.

The fairness problem is ancient and dates back at least to classical Greece. It has been treated recently by mathematicians who typically are concerned with the existence of a "fair division" of a nonuniform object among n persons; that is, a division with the property that each party thinks he is getting at least $1/n$th of the value of the object. (See, for example, Lester Dubins and Edwin Spanier, Harold Kuhn, and Hugo Steinhaus.) This is not the approach we will take, since we will assume a world of homogeneous infinitely divisible goods in which the mathematical fair division problem becomes trivial.

The concept of fairness has also been treated extensively by philosophers. The most recent philosophical approach is that of John Rawls, who argues at length for a social contract theory of justice: a society which maximizes the welfare of its worst off members is most just and that is the sort of society people will, from an initial position of ignorance about their endowments and interests, contract to enter. Rawls' approach has been extended to a theory of taxation by Edmund Phelps. Again, Rawlsian fairness, or "justice," is not the fairness we are interested in; we do not assume a precontractual state of ignorance, we do assume that knowledge of wealth and tastes is given. In fact, knowledge about one's own and others' bundles of goods is crucial in our discussion.

What then is our notion of fairness? It is fairness in the sense of *non-envy*. A completely fair social state is one in which no citizen would prefer what another has to what he himself has; a relatively fair social state is one in which few citizens would prefer what others have to what they themselves have; a totally unfair state is one in which every citizen finds his position to be inferior to that of everyone else. This concept of fairness is appealing because it only depends, like other economic concepts, on individual tastes and endowments.

Fairness in the non-envy sense has been discussed in several recent papers by economists. Serge Christophe Kolm considers allocative fairness, and shows that there exist allocations which are both completely fair and efficient.[1] David Schmeidler and Karl Vind define fair *trades* as

* Brown University and Center for Operations Research and Econometrics, Catholic University of Louvain, respectively.

[1] In a recent paper Hal Varian extends this sort of analysis. Also Richard Zeckhauser discusses one of Kolm's concerns at some length: what does a fairness-minded planner do when there is an unfair distribution of nontransferrable goods, like I.Q.?

trades which exclude envy (a transaction between a price discriminating monopolist and two customers—each getting a different price—is an "unfair" trade in this sense). In both the Kolm and the Schmeidler-Vind papers, things are either completely fair (or totally devoid of envy) or not. Neither paper considers the problem of establishing an index of fairness and maximizing it, as we will below.

In this paper we will first discuss the question of the preservation of complete fairness. We will show that such fairness is *not* preserved by competitive equilibrium trades, by trades to the core, or *even* by fair trades. In short, perfect fairness is a delicate condition. Whether this says something about our definitions, or whether it says something about the real world, we leave to the reader to decide. Our own feeling is that it says something about both.

Then we will define three social measures of envy. One is an ordinal-utility counting measure which sums up the instances of envy inherent in an allocation. The second and third are weighted sums of cardinal-utility individual envy measures. Finally we will characterize the solutions of some fairness maximization (or envy minimization), problems in an economy in which all traders have identical utility functions.

I. The Model

Consider an n-trader, m-good economy. We will assume that each of the m commodities is homogeneous and infinitely divisible. An *allocation* $x=(x_1, x_2, \dots, x_n)$ is an n vector of nonnegative m vectors, i.e., a point in R_+^{nm}, whose ith component, x_i, is the bundle of goods assigned to trader i under x. We will let ω be a fixed initial allocation, we assume that $\sum_{i=1}^n \omega_i > 0$.[2] We also define $A(\omega)$ to be the set of allocations which are feasible in our economy. That is, $A(\omega) \equiv \{x\colon x \geqq 0 \text{ and } \sum_{i=1}^n x_i = \sum_{i=1}^n \omega_i\}$.

We will assume that every trader has preferences which can be represented by a continuous utility function $u_i(x_i)$ which maps R^m into R.

We will assume that the utility functions of all traders are strictly quasi concave: If x_i and y_i are distinct nonnegative m vectors, $0<\lambda<1$, and $u_i(x_i) \geqq u_i(y_i)$, then $u_i(\lambda x_i + (1-\lambda) y_i) > u_i(y_i)$. We will also assume that all u_i are strictly monotonic.

Pareto optimal (or *efficient*), *core*, and *competitive equilibrium* allocations are defined in the usual ways. Consider an allocation x. If a subset S of traders can redistribute its own resources in a way which makes all of its members at least as well off as x makes them, and makes some of them better off, we say that S can block x. Formally, if there are bundles s_i, for all i in S, such that $\sum_{i \text{ in } S} s_i = \sum_{i \text{ in } S} \omega_i$ $u_i(s_i) \geqq u_i(x_i)$ for all i in S, and $u_j(s_j) > u_j(x_j)$ for some j in S, then S *blocks* x. An allocation is in the core if no group of traders can block it. An allocation is Pareto optimal if it cannot be blocked by the whole set of traders. If p is a vector of prices and $\hat{x}$ is an allocation, and if $\hat{x}_i$ maximizes $u_i(x_i)$ subject to $p \cdot x_i \leqq p \cdot \omega_i$ for all i, then we say that $(p, \hat{x})$ is a competitive equilibrium, and that $\hat{x}$ is a competitive equilibrium allocation.

Following Kolm and Schmeidler and Vind, we will say that an allocation x is *fair* if for every pair of traders $\{i, j\}$, $u_i(x_i) \geqq u_i(x_j)$. If $t=(t_1, t_2, \dots, t_n)$ is a vector of m vectors satisfying $x_i + t_i \geqq 0$ for all i and $\sum_{i=1}^n t_i = 0$, we will say that t is a *feasible trade from* x. We will say that t is a *fair trade from* x if it is feasible and if

[2] We use the following vector inequality notation: If x and y are k dimensional vectors, $x \geqq y$ means $x_i \geqq y_i$ for $i=1, \dots, k$; $x \geq y$ means $x_i \geqq y_i$ for $i=1, \dots, k$, and $x_j > y_j$ for some j; $x > y$ means $x_i > y_i$ for $i=1, \dots, k$.

for every pair $\{i, j\}$ of traders, $u_i(x_i+t_i) \geqq u_i(x_i+t_j)$ whenever $x_i+t_j \geqq 0$.

According to our definitions, a fair allocation is an allocation with the property that no trader would prefer another's bundle of goods to his own, and a fair trade is a trade with the property that no trader would prefer another's *exchange* to his own, providing that he could have made it.

Now we will define $C(x) \equiv$ the number of pairs $\{i, j\}$ for whom $u_i(x_i) < u_i(x_j)$. That is, $C(x)$ counts the number of instances of envy associated with the allocation x. Clearly $C(x)=0$ if and only if x is fair and $C(x)$ attains its maximum of n^2-n when every trader is envious of every other trader. When we maximize fairness in the sense of $C(x)$, we will be maximizing $-C(x)$.

Our second and third, nondiscrete, measures of envy presume that individuals have cardinal utility functions, so that utility sums and differences are meaningful for each trader.

Let us define

$$e_i(x) \equiv \sum_{j=1}^{n} [u_i(x_j) - u_i(x_i)]$$

The function $e_i(x)$ measures i's total envy by adding up his envy (positive or negative) of every other trader. We can define a vector of envies as follows:

$$e(x) \equiv (e_1(x), e_2(x), \ldots, e_n(x))$$

The vector $e(x)$ can be treated in a manner analogous to the usual economic treatment of utility vectors; for example, it is possible to find allocations undominated in envy just as it is possible to find allocations undominated in utility. We might remark that if x is a fair allocation, then $e(x) \leqq 0$, but not vice versa.

The measure $e_i(x)$ has the property that a man who isn't on the bottom rung of the economic ladder is compensated (in envy) by the misfortune of those below him. Such compensation may not seem entirely natural, and for this reason we will define a third measure of envy without this feature. Let

$$\overset{*}{e}_i(x) \equiv \sum_{j:u_i(x_j) \geqq u_i(x_i)} [u_i(x_j) - u_i(x_i)]$$

Also, let

$$\overset{*}{e}(x) \equiv (\overset{*}{e}_1(x), \overset{*}{e}_2(x), \ldots, \overset{*}{e}_n(x))$$

Now we can remark that x is a fair allocation if and only if $e^*(x) = (0, 0, \ldots, 0)$.

The two vectors $e(x)$ and $e^*(x)$ are not comparable to the scalar $C(x)$, so we will define two more social envy (or unfairness) measures as follows:

If $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a set of positive weights, we will let the *total social envy* (given α) for an allocation x be

$$E(\alpha, x) \equiv \sum_{i=1}^{n} \alpha_i e_i(x)$$

Similarly, we will let the *total social envy** (given α) for an allocation x be

$$\overset{*}{E}(\alpha, x) \equiv \sum_{i=1}^{n} \alpha_i \overset{*}{e}_i(x)$$

Now for any $\alpha > 0$, $E(\alpha, x) \leqq 0$ if x is fair, and $E^*(\alpha, x) = 0$ if and only if x is fair. When we maximize fairness in the sense of $E(\alpha, x)$ or $E^*(\alpha, x)$, we will be maximizing $-E(\alpha, x)$ or $-E^*(\alpha, x)$.

II. The Delicate Nature of Allocative Fairness

Partly for our immediate gratification, and partly to motivate the discussion of fairness maximization which will follow, we will now show that a number of reasonable looking, useful, sensible, and comforting conjectures about fairness are false.

In this section we are particularly concerned with the preservation of allocative fairness. A goal of social policy ought to have some "stability" properties; if the goal is "unstable" in some sense, the policy

maker's job is much bigger than if it is "stable": he must not only institute a change, he must also remain around forever to prevent backsliding. Pareto optimality has an obvious stability property since deviations from it create self-corrective incentives. This is not the case for fairness, however.

Our first two examples are motivated by a theorem of Kolm which says that (under general conditions) there exist allocations which are simultaneously fair and efficient. The proof of the theorem uses the fact that a trade from the equal allocation (which assigns every trader an identical bundle of goods) to a competitive equilibrium preserves fairness. The reason for this is transparent, for if the economy starts at the equal allocation and x is a competitive allocation based on it, then x_j must be in the ith trader's budget set for every pair $\{i, j\}$, and so envy (that is, an inequality of the form $u_i(x_j) > u_i(x_i)$) contradicts utility maximization. But what if we start at an arbitrary fair allocation and make a competitive equilibrium trade? Do we end at a competitive equilibrium allocation which is fair?

The following Edgeworth box diagram shows that we need not. In Figure 1, i_1 and i_2 are two of trader I's indifference curves; j_1 and j_2 are two of trader J's; $\omega = (\omega_i, \omega_j)$ is the initial allocation; $\omega^{-1} \equiv (\omega_j, \omega_i)$ is the allocation which switches the bundles between i and j. Now the allocation $x = (x_i, x_j)$ is a competitive allocation (from ω), but it is *not* fair, because $x^{-1} \equiv (x_j, x_i)$ lies above the indifference curve labelled i_2 which means that trader I envies trader J at x.

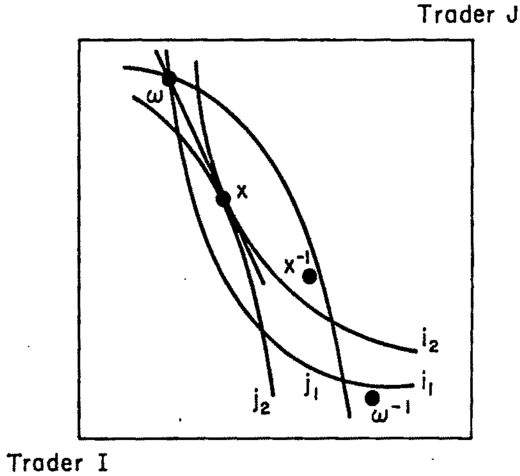


FIGURE 1

We've shown that competitive equilibrium trades may destroy fairness. Moreover, a trade from the equal allocation (surely the fairest of the fair) to the core may also destroy fairness. An example of this perverse result is illustrated in Table 1. In this three-person three-good economy, the initial allocation is the equal allocation. Some examination will convince the reader that x is in the core: no subset of the three traders could, by an internal redistribution of its initial holdings, make all of its members at least as well off, and some better off, than x makes them. However, $u_2(x_1) = 20/3 > 6 = u_2(x_2)$, so x is not fair. Therefore, barter exchange is apt to destroy fairness, even from a starting point of complete equality.

Our final example shows that fair trades themselves may destroy (allocative) fairness. In this sense the approaches of Schmeidler and Vind and Kolm are mu-

TABLE 1

	Utility Functions	ω_i	$u_i(\omega_i)$	x_i	$u_i(x_i)$
Trader 1	$u_1(x_1) = 3x_{11} + 2x_{12} + x_{13}$	(1, 1, 1)	6	(3, 2/3, 0)	10 1/3
Trader 2	$u_2(x_2) = 2x_{21} + x_{22} + 3x_{23}$	(1, 1, 1)	6	(0, 0, 2)	6
Trader 3	$u_3(x_3) = x_{31} + 3x_{32} + 2x_{33}$	(1, 1, 1)	6	(0, 7/3, 1)	9

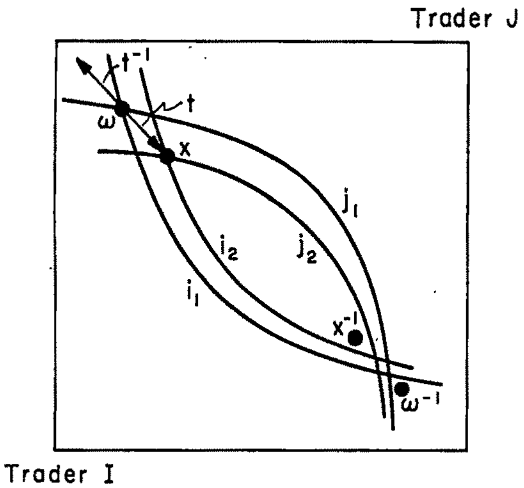


FIGURE 2

tually inconsistent. Moreover, the example best illustrates the delicacy of fairness; it shows that the preservation of complete allocative fairness requires more than just a monitoring of the fairness of the moves; it requires a constant monitoring of the *results* of the moves. Consider the Edgeworth box diagram of Figure 2.

Here we again have an economy of two traders and two goods: again i_1 and i_2 are two of trader I's indifference curves; j_1 and j_2 are two of trader J's; $\omega=(\omega_i, \omega_j)$ is the initial allocation; $t=(t_i, t_j)$ is a feasible trade, and $x=\omega+t$; ω^{-1} is again (ω_j, ω_i) and $x^{-1}\equiv(x_j, x_i)$. Also, $t^{-1}\equiv(t_j, t_i)$. Since both i_1 and j_1 pass "above" ω^{-1}, ω is *fair*. Since t makes I better off while t^{-1} would make him worse off, and a symmetrical argument applies to trader J, t is a *fair trade from* ω. But i_2 passes "under" x^{-1} as does j_2 and both I and J are envious at x. Therefore, x is *not* fair. A fair trade from a fair allocation can result in an unfair allocation.

When it's defined as the total absence of envy, fairness is a fragile condition. It is apt to disappear if people engage in trade for private benefit.

Having said this much, we will move on to less demanding criteria of fairness than the total absence of envy. Now let's measure the extent of envy, and consider the problem of minimizing it, without making anyone worse off. We are interested then in the qualitative implications, if there are any, of envy minimization.

III. When $C(x)$ Measures Envy

In this section we will analyze the problem of maximizing fairness in the sense of minimizing $C(x)$. The unconstrained problem is of course trivial, since it is solved by the equal allocation, among others. The interesting approach is to minimize $C(x)$ subject to a constraint, and the most obvious constraint is the requirement that no trader be made worse off by a fairness-increasing move. This is the natural constraint of voluntariness, natural because in a free society there is usually a governmental predisposition toward Pareto moves.

We are concerned, then, with minimizing $C(x)$ subject to $u_i(x_i)\geqq u_i(\omega_i)$, for all i. The problem clearly has a solution since $0\leqq C(x)\leqq n^2-n$ and C takes on only integer values. However, it cannot generally be solved by standard methods, so we will confine our analysis to a special case. We will assume that every trader has the *same* strictly quasi-concave, monotonic utility function u. We will also assume that u is homothetic: for $\lambda\geqq 0$, $u(\lambda x)=\phi(\lambda)u(x)$, where $\phi(\lambda)$ is some monotonic function of λ. Under these conditions it is possible to "reduce" the economy to one in which: 1) every trader has a bundle $\hat{x}_i$ which is proportional to $\sum_{i=1}^n \omega_i$; 2) there is a "social surplus" bundle $L\equiv\sum_{i=1}^n \omega_i-\sum_{i=1}^n \hat{x}_i$; 3) $\hat{x}\equiv(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$ is efficient in the economy with total resources $\sum_{i=1}^n \omega_i-L=\sum_{i=1}^n \hat{x}_i$; and 4) $u(\hat{x}_i)=u(\omega_i)$, for all i. Since all the $\hat{x}_i$ and L are proportional to ω, we can simply define $\sum_{i=1}^n \omega_i$ to be one unit of one *composite* good.

Now let us imagine that the economy

has been reduced, that is, that the allocation $(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n, L)$ satisfying 1)–4) above has been established, and that an enlightened ruler wants to distribute L in a way which will maximize fairness in the sense of minimizing $C(x)$. He can do so by distributing proportional bundles from the social surplus bundle L since the distribution of proportional bundles will preserve efficiency. Therefore, we can restrict our attention to the one-composite-good case. Let ΔL_i be the "fairness grant" of the composite good going to the ith individual. The ruler's problem is to choose $(\Delta L_1, \Delta L_2, \ldots, \Delta L_n) \geqq 0$ so that $\sum \Delta L_i \leqq L$, and so that the number of pairs $\{i, j\}$ for whom $u(\hat{x}_i + \Delta L_i) > u(\hat{x}_j + \Delta L_j)$ is minimized. By the monotonicity assumption, the inequality $u(\hat{x}_i + \Delta L_i) > u(\hat{x}_j + \Delta L_j)$ can be replaced by $\hat{x}_i + \Delta L_i > \hat{x}_j + \Delta L_j$.

We will partition the traders in the economy into $h \leqq n$ classes, $S_1, S_2, \ldots, S_h$, by putting traders with equal x_is into the same class. Let us suppose $\hat{x}_{i_1}$ is the $\hat{x}_i$ associated with class S_1, $\hat{x}_{i_2}$ is the $\hat{x}_i$ associated with S_2, and so on, and without loss of generality, we'll assume the classes are numbered from richest to poorest: $\hat{x}_{i_1} > \hat{x}_{i_2} > \ldots$. Now define $\delta_2 \equiv \hat{x}_{i_1} - \hat{x}_{i_2}$, the difference between the wealth of members of class S_1 and members of class S_2, $\delta_3 \equiv \hat{x}_{i_2} - \hat{x}_{i_3}, \ldots, \delta_h \equiv \hat{x}_{i_{h-1}} - \hat{x}_{i_h}$. Finally, suppose $n_1 \equiv$ the number of members of S_1, $n_2 \equiv$ the number of $S_2, \ldots, n_h \equiv$ the number of members of S_h.

The only way to eliminate instances of envy (without making anyone worse off) is to move groups of traders from lower classes to higher classes. It is clear that any total migration upward can be represented as a vector of one-step upward moves. Therefore, any movement upward can be represented by a vector $(k_2, k_3, \ldots, k_h)$, where $k_r \equiv$ the number of individuals who move from S_r to S_{r-1}. Since we cannot have negative numbers of individuals in any class, we must have

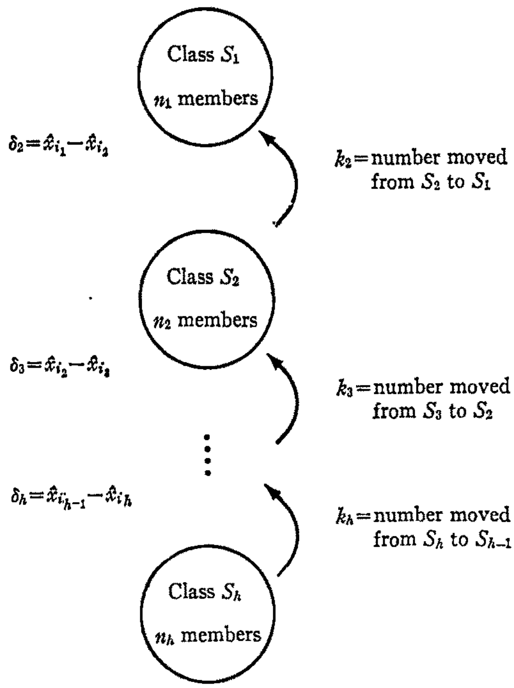


FIGURE 3

$n_r - k_r + k_{r+1} \geqq 0$ for $r = 2, \ldots, h-1$, and $n_h - k_h \geqq 0$. The total "cost" to the dictator of a movement represented by $(k_2, \ldots, k_h)$ is given by $\sum_{r=2}^{h} \delta_r k_r$. The process is illustrated in Figure 3.

It's easy to see that the number of instances of envy associated with the allocation $\hat{x}$ is given by

$$C(\hat{x}) = \sum_{i<j} n_i n_j, \text{ where } \quad 1 \leqq i \leqq j \leqq h$$

If the social surplus is disbursed in a way which brings about a vector of upward movements $(k_2, k_3, \ldots, k_n)$, the number of instances of envy becomes

$$\begin{aligned}(n_2 - k_2 + k_3)&(n_1 + k_2) + (n_3 - k_3 + k_4) \\ &\cdot (n_2 - k_2 + k_3 + n_1 + k_2) + \ldots \\ &= \sum_{i<j} n_i n_j + \sum_{r=2}^{h} k_r(n_r - n_{r-1}) \\ &\quad + \sum_{r=2}^{h} k_r(k_{r-1} - k_r)\end{aligned}$$

where we define $k_1 \equiv 0$.

We can finally reformulate the benevolent ruler's fairness maximization problem as follows: Having extracted the social surplus from the economy and brought it to an allocation $\hat{x}$, he wants to disburse that surplus, to establish a new allocation, say x, in a way that maximizes the reduction in envy (or minimizes the increase) brought about by his disbursement. That increase is given by

$$C(x) - C(\hat{x}) = \sum_{r=2}^{h} k_r(n_r - n_{r-1}) + \sum_{r=2}^{h} k_r(k_{r-1} - k_r)$$

so he wants to choose $(k_2, \ldots, k_h) \geqq 0$ to minimize $\sum_{r=2}^{h} k_r(n_r - n_{r-1}) + \sum_{r=2}^{h} k_r(k_{r-1} - k_r)$ subject to $k_1 = 0$, $n_r - k_r + k_{r+1} \geqq 0$ for $r = 1, \ldots, h-1$, $n_h - k_h \geqq 0$, and $\sum_{r=2}^{h} \delta_r k_r \leqq L$.

It is worth noting that a solution for this problem might be found for which

$$\sum_{r=2}^{h} \delta_r k_r < L$$

In such a case there is a leftover $L - \sum_{r=2}^{h} \delta_r k_r$ which cannot be used to reduce envy. This leftover can clearly be divided up and distributed in such a way that no one is moved out of his own class; the (fairness-maximizing) allocation which results is then efficient in the original (unreduced) economy.

The analysis of this fairness maximization problem is straightforward, providing we ignore the implicit integer constraints (only whole persons can be moved) on the k_is. First, we remark that the quadratic part of the objective function, $\sum_{r=2}^{h} k_r(k_{r-1} - k_r)$, is concave.[3] Therefore, the objective function is concave and the problem is one of minimizing a concave function on a closed and bounded convex set. It follows that it will have a solution at an extreme point of that feasible set. Therefore, the solution can be characterized in one of the following ways. (For notational simplicity we define $k_{h+1} \equiv 0$.)

PROPOSITION 1: *The solution to the fairness maximization (in the sense of $C(x)$ minimization) problem satisfies one of the following sets of conditions:*

$$(1) \qquad \sum_{r=2}^{h} \delta_r k_r < L$$

and for each $i = 1, 2, \ldots, h$, either $k_i = 0$ or $k_i = n_i + k_{i+1}$.

$$(2) \qquad \sum_{r=2}^{h} \delta_r k_r = L$$

and for all i but at most one, either $k_i = 0$ or $k_i = n_i + k_{i+1}$.

Let us interpret these conditions: (1) says that every class is either eliminated or has *no* out-migrants whatsoever. There are two degenerate subsolutions:

$$(1') \qquad k_r = n_r + k_{r+1}, \qquad \text{for } r = 2, \ldots, h$$

Now everyone moves up to the first class, which means that we must have had

---

[3] If we let

$$M \equiv \begin{vmatrix} -1 & 1/2 & 0 & 0 & \ldots \\ 1/2 & -1 & 1/2 & 0 & \ldots \\ 0 & 1/2 & -1 & 1/2 & \ldots \\ \vdots & & & & \\ 0 & 0 & \ldots & 1/2 & -1 \end{vmatrix}$$

then

$$\sum_{r=2}^{h} k_r(k_{r-1} - k_r) = (k_2, \ldots, k_h) M \begin{bmatrix} k_2 \\ \vdots \\ k_h \end{bmatrix}$$

and the quadratic form is concave if M is negative definite (see, e.g., George Hadley), that is, if $-M$ is positive definite. Now $-M$ is positive definite if it has a positive dominant diagonal (see, e.g., Hukukane Nikaido, pp. 385–87), and it is a simple exercise, which we will not perform here, to show that $-M$ does in fact have a positive dominant diagonal.

$$\sum_{r=2}^{h} \delta_r k_r = \sum_{r=2}^{h} \delta_r \left[\sum_{i=r}^{h} n_i \right] < L$$

to begin with: the social surplus was large enough to establish complete equality.

$$(1'') \qquad k_2 = k_3 = \ldots = k_n = 0$$

Now no one moves. As an example, consider the case where $n_1=100$, $n_2=1000$, $\delta_2=1$, and $L=10$. Any choice of k_2 greater than zero (and necessarily smaller than 10), will cause envy to increase. It is clear that a necessary condition for the no-movement solution is that the class sizes be pyramidal; that is, $n_1<n_2< \ldots <n_h$.

For solutions in category (2), every class but one is either disappearing, or absorbing. The degenerate solution is again given by

$$(2') \qquad k_r = n_r + k_{r+1}, \qquad \text{for } r = 2, \ldots, h$$

Everyone moves up to the first class.

In *no* case do we have to be concerned with the possibility that some but not all individuals in class i move and some but not all individuals in class j move. Moreover, there is *no* presumption that the fairness disbursement need go first to the poorest classes.

Let us observe that the solution to the fairness maximization problem could be found by an exhaustive search of the extreme points of the feasible set. Is there a marginal algorithm which will also find it? The marginal gain in fairness (or reduction in envy) which results from the movement of one member of class j to class $j-1$ is given by

$$-n_j n_{j-1} + (n_j - 1)(n_{j-1} + 1) = n_j - 1 - n_{j-1}$$

Might we then not start out by moving that class j for which $(n_j-1-n_{j-1})/\delta_j$ is largest? Unfortunately, such local rules are unsatisfactory, because our problem is one of minimizing a concave, rather than convex, function over a convex set. For illustration, suppose

$$n_1 = 100 \quad n_2 = 50 \quad n_3 = 30 \quad n_4 = 6$$
$$\delta_2 = 2 \quad \delta_3 = 4 \quad \delta_4 = 1 \qquad L = 6$$

The marginal benefit-cost ratio is highest for a movement of a member of class 2 into class 1. Inspection reveals, however, that the maximum envy reduction is obtained by moving all the members of class 4 into class 3.

Let's summarize the above discussion. The solution to $C(x)$ minimization problem is an extreme solution, extreme in the sense that for all but at most one class, classes must be moved in their entirety, or they must be absorbing. Moreover, it is the *classes* that are crucial, since the strengths of individual feelings do not appear in $C(x)$. And, finally, there is no reason to believe that the poorest classes will be moved first.

What happens to this analysis if we relax our assumption that everyone has the same utility function u? If there is a *single* good in the economy, and if each person's utility function u_i increases monotonically in it, the analysis of the disbursement of a social surplus goes through exactly as it does above. (Where that surplus comes from, however, becomes problematical, since there is no inefficiency to begin with.) However, if there is *more* than one good, we run into difficulty. Given an allocation x, we can define a non-envy relation R on the set of individuals in the economy by saying iRj ("i doesn't envy j") whenever $u_i(x_i) \geqq u_i(x_j)$. Now if x is efficient, it can be shown that the relation R is complete and acyclic, and we can, therefore, given any subset of individuals, identify persons who envy no one in that subset (see, for example, Amartya Sen). But R may *not* be transitive, and it may therefore be impossible to define an envy class structure, as we've done above. The solution to the $C(x)$ minimization problem

still exists, of course, but our discussion of how to find it becomes irrelevant.

IV. When $e_i(x)$ and $e^*(x)$ Measure Envy

Now we will consider the problem of maximizing fairness in the sense of minimizing our cardinal utility envy measures $E(\alpha, x)$ and $E^*(\alpha, x)$. We will again require that no one be made worse off, and we will again make the strong assumption that everyone has the same utility function $u=u_1=u_2=\ldots=u_n$. Rather than discussing the reduction of a many-goods economy to an efficient one-composite-good economy, we will presume at the outset that there is just one good. We will assume throughout this section that u is continuous and has a continuous first derivative u'. We will also suppose that u is nondecreasing and u' is nonincreasing: If $x_i \geqq x_j$, then $u(x_i) \geqq u(x_j)$ and $u'(x_i) \leqq u'(x_j)$. If we interpret x_i as i's income, then the marginal utility of income is nonnegative, and nonincreasing. We are now starting with a distribution of goods $(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n, L)$, where L is the social surplus which will be distributed by the benevolent ruler.

The constrained minimization of $E(\alpha, x)$ is trivial under these circumstances, for any set of weights $\alpha=(\alpha_1, \ldots, \alpha_n) \geqq 0$. For under the above assumptions,

$$E(\alpha, x) = \sum_{i=1}^{n} \alpha_i e_i(x)$$

$$= \sum_{i=1}^{n} \alpha_i \left[\sum_{j=1}^{n} (u(x_j) - u(x_i)) \right]$$

$$= \sum_{i=1}^{n} (A - n\alpha_i) u(x_i),$$

where $A \equiv \sum_{i=1}^{n} \alpha_i$

To minimize $E(\alpha, x)$ the ruler can distribute all of the social surplus to an individual for whom α_i is largest, that is, an individual whose envy is given the greatest weight in the calculation of total social envy. If all individuals are weighted equally, then any and all allocations are $E(\alpha, x)$ minimizing, since $E((1/n, 1/n, \ldots, 1/n), x) \equiv 0$ for all x.

This rather amoral result stems from the fact that the measure $e_i(x)$ counts negative envy, or the psychic compensation i gets from the misfortune of those below him, as well as his jealousy of those above. When all the u_i are identical, and $\alpha=(1/n, \ldots, 1/n)$, the jealousies and the psychic compensations cancel out. This will clearly not be the case for $e^*(x)$ and, of course, $E^*(\alpha, x)$, since $e^*(x)$ does not share $e_i(x)$'s I-feel-better-when-others-are-hurt property.

Let us then turn to the minimization of $E^*(\alpha, x)$ via a distribution of the social surplus L. A simple example will illustrate the nature of this problem.

Suppose our economy has three members: $u(x) \equiv x$; $\hat{x}_1=10$; $\hat{x}_2=0$, and $\hat{x}_3=0$. Trader 1 is the rich man, and 2 and 3 are equally poor. Assume that $L=1$, $\alpha_1=1/10$, $\alpha_2=2/10$, and $\alpha_3=7/10$. Note that the envy weights are assigned to *persons*, not to positions in the hierarchy, so $\alpha_2 \neq \alpha_3$ although both 2 and 3 are equally poor. In this sense, $E^*(\alpha, x)$ is not "neutral" between persons or blind to individual identification. In our example the fairness maximizing disbursement of L is simply a grant of 1 to individual 3. Giving equal shares to 2 and 3 is not the way to minimize $E^*(\alpha, x)$. Moreover, the result would hold even if $\hat{x}_3$ were *greater* than zero. Therefore, fairness maximization (in the sense of $E^*(\alpha, x)$ minimization) may not only create instances of envy, it may also involve a policy of grants to the (relatively) rich, and one of benign neglect toward the poor.

However, this possibility disappears when we force some degree of neutrality on $E^*(\alpha, x)$. There are (at least) two ways to do so: The first is to assign weights to positions in the hierarchy rather than persons, and to assume that the envy of the

poorer man is always weighted at least as heavily as the envy of the richer man. The second and simpler way is to assume that $\alpha_1=\alpha_2= \ldots =\alpha_n$. Let's now take this latter approach. If $\alpha=(1/n, \ldots, 1/n)$, then

$$E^*(\alpha, x) = \sum_{i=1}^{n} \alpha_i \sum_{j:u(x_j)\geqq u(x_i)} [u(x_j) - u(x_i)]$$
$$= \frac{1}{n}\sum_{i=1}^{n} \sum_{j:u(x_j)\geqq u(x_i)} [u(x_j) - u(x_i)]$$

We will assume, without loss of generality, that our individuals are indexed so that $\hat{x}_1 \geqq \hat{x}_2 \geqq \ldots \geqq \hat{x}_n$. Since $\alpha_i=1/n$ for all i, and $u_i=u$ for all i, $E^*(\alpha, x)$ can clearly be minimized in a way which does not affect the rank order of wealth. In other words, it can be minimized by an allocation x with the property that

$$(3) \qquad x_1 \geqq x_2 \geqq \ldots \geqq x_n$$

As long as (3) holds, however, we have

$$E^*\left(\left(\frac{1}{n}, \ldots, \frac{1}{n}\right), x\right) = \frac{1}{n}\sum_{i=1}^{n} \sum_{j\leqq i} [u(x_j) - u(x_i)]$$

which gives, after some manipulation,

$$E^*\left(\left(\frac{1}{n}, \ldots, \frac{1}{n}\right), x\right) = \frac{n+1}{n}\sum_{i=1}^{n} u(x_i) - \frac{2}{n}\sum_{i=1}^{n} iu(x_i)$$

It follows that an incremental increase in x_i increases envy by an amount

$$\frac{n+1}{n} u'(x_i)dx_i - \frac{2}{n} iu'(x_i)dx_i = \left(\frac{n+1-2i}{n}\right) u'(x_i)dx_i$$

as long as (3) holds. Since we have assumed u' is nonnegative and nonincreasing, this term is smallest when $i=n$, so it's always best to bestow upon the humblest. When there are ties for last place (when, for example, $x_n=x_{n-1}$, etc.), equal shares must be given to all the poorest in order to preserve (3). This argument establishes:

PROPOSITION 2: *If $\alpha=(1/n, \ldots, 1/n)$, and marginal utility is nonnegative and nonincreasing, then the fairness maximization (or E^* minimization) problem is solved by a policy of pushing the poorest up from the bottom: The ruler gives to trader n until $x_n=x_{n-1}$, then he gives to n and $n-1$ until $x_n=x_{n-1}=\hat{x}_{n-2}$, then he gives to n, $n-1$, and $n-2$ until $x_n=x_{n-1}=x_{n-2}=\hat{x}_{n-3}$, and so on.*

Proposition 2 is the morally classical result, and seems almost obvious on its surface. We might remark that it is analogous to the utilitarian argument that social welfare (defined as a sum of identical individual utility functions) is maximized through an equal distribution of income. Like that argument, it depends on the assumptions of (i) identical and therefore comparable utility functions, and (ii) decreasing marginal utility of income. Both assumptions are perhaps more plausible from the philosopher's viewpoint than from the economist's.

V. Conclusion

This paper has three major points. The first is that standard voluntary economic transactions have little apparent connection with the fairness, or lack of fairness, of allocations. In general, even if economic transactions are fair in themselves, like trades to competitive equilibria, they can be expected neither to establish nor to preserve allocative fairness. Fairness, unlike efficiency, has no automatic enforcers.

Second, an envy measure which simply counts instances of envy imposes certain types of solutions on a benevolent dictator's constrained fairness maximization problem. The discontinuity of the counting measure forces the dictator to look at

classes, rather than individuals, since the measure is a function of numbers in classes, rather than intensities of individual envies. Moreover, the problem is such that it will have extreme solutions: with one possible exception, classes will be moved up in their entireties, or not moved up at all. And, finally, there is no assurance that it is the poorest classes which will be moved.

Third, envy measures which assume cardinal utility, or which depend on intensities of individual envies, lead the benevolent dictator down different paths. If an envy measure includes psychic compensation that the rich receive from the poverty of those poorer than themselves, the fairness optimizing policy may be to do nothing. If the rich are assumed to get no satisfaction from the poverty of the poor, an enlightened ruler may, under certain conditions, maximize fairness by giving society's excess to the poorest. This is, of course, the most intuitive solution, but it is a solution which depends on rather stringent assumptions.

REFERENCES

L. E. Dubins and E. H. Spanier, "How to Cut a Cake Fairly," *Amer. Math. Mon.*, Jan. 1961, *68*, 1–17.

G. Hadley, *Nonlinear and Dynamic Programming*, Reading, Mass. 1964.

S. C. Kolm, "Justice et Équité," mimeo., Centre d'Études Prospectives d'Économie Mathématique Appliquées à la Planification, Paris 1971.

H. W. Kuhn, "On Games of Fair Division," in M. Shubik, ed., *Essays in Mathematical Economics in Honor of Oskar Morgenstern*, Princeton 1967, 29–37.

H. Nikaido, *Convex Structures and Economic Theory*, New York 1968.

E. S. Phelps, "Taxation of Wage Income for Economic Justice," *Quart. J. Econ.*, Aug. 1973, *87*, 331–54.

J. Rawls, *A Theory of Justice*, Cambridge 1971.

D. Schmeidler and K. Vind, "Fair Net Trades," *Econometrica*, July 1972, *40*, 637–42.

A. K. Sen, *Collective Choice and Social Welfare*, San Francisco 1970.

H. Steinhaus, "Sur la Division Pragmatique," *Econometrica*, July 1949, Suppl., *17*, 315–19.

H. R. Varian, "Equity, Envy, and Efficiency," working pap. in econ. theory and econometrics, no. IP-172, Univ. California, Berkeley 1972.

R. Zeckhauser, "Risk Spreading and Distribution," pap. no. 10, Kennedy School of Government Public Policy Program, Cambridge 1972.

Investment Decisions Under Uncertainty: The "Irreversibility Effect"

By Claude Henry*

What is meant here by an "irreversible decision?" A decision is considered irreversible if it significantly reduces for a long time the variety of choices that would be possible in the future.

Suppose for instance that we in 1974 must decide whether the cathedral of Notre-Dame is to be preserved or to be demolished and replaced by a parking lot. A decision to preserve Notre-Dame is not irreversible since, if adopted in 1974, it leaves open the possibility of further choices between Notre-Dame and the parking lot (or other alternatives). However, if the parking lot is built in 1974, we will never again be in a position to choose to keep Notre-Dame; this certainly is an irreversible decision.

After ploughing his field, a farmer considers what kind of crop to plant now. Whatever decision he adopts is not irreversible, since he will later have to decide what to plant again in this field; the set of possibilities offered to him will not have been modified by the decision previously taken. On the other hand, if he decides to hew down a forest of full-grown oaks and bring more land into cultivation, he makes an irreversible decision. Generally, it is the different degrees of irreversibility associated with various possible decisions which are of interest: to build a power station where coal can be burned as well as oil is a "less irreversible" decision than to build an equally powerful station where only oil can be used as a fuel.

Our concern with irreversibilities[1] is actually related to the following problem: a new circumferential highway is now being planned around Paris as a direct connection between the various suburbs located ten kilometers beyond the city limits. It may cut through public forests (Versailles, Malmaison, etc.), ancient royal estates, that form a first green belt west of Paris. Will they be spared or will they be partially destroyed? The "Ministère de l'Equipement," responsible for highway planning, and the "Ministère de l'Agriculture," responsible for the management of public forests, are involved in a cost-benefit analysis on this point.

According to a rule systematically adopted at the "Ministère de l'Equipement," every random return or cost appearing in the problem is replaced by its expected value before application of any decision criterion (see, for example, G. Dreyfus). The initial random problem is thus replaced by an associated riskless problem, i.e., a (supposedly) "equivalent certainty case" in the sense of Herbert Simon and Henri Theil. Our approach here[2] is to show that irreversibility makes it impossible to draw the conclusions of Simon and Theil even if all the other assumptions of their models—including the quadratic payoff function—are satisfied.[3] Irreversibility even prevents the use of the concept Edmond Malinvaud calls "first-order certainty equivalence."

In fact using the information structure

* Centre National de la Recherche Scientifique and Ecole Polytechnique, Paris. This research has been supported by the C.N.R.S. under contract A.T.P. no. 62990 1. I am grateful to J. H. Drèze, G. M. Heal, A. P. Kirman, and E. Malinvaud, and an unknown referee for helpful comments, but of course, retain entire responsiblity for errors. Special thanks are due to A. C. Duyck for his help in computing the numerical examples.

[1] For other examples of irreversibilities, see P. W. Barkley and D. W. Seckler, pp. 149–55, and Anthony Fisher, John Krutilla, and Charles Cicchetti.

[2] In Kenneth Arrrow and Fisher, the authors are making a similar point in a different framework.

[3] In the Appendix, a counterexample is presented which shows that, if we introduce irreversibility in Theil's model without changing any other of his assumptions, his result is no longer valid, i.e., the certainty equivalence does not hold, even for the first period.

defined in the next paragraph I will prove that, by replacing the initial random problem, even a risk-neutral[4] decision maker facing a binary alternative is led to adopt an irreversible decision more often than he should. Of course it is not surprising that, in general, replacing random variables by their expectations will not lead to the appropriate decisions. Our point is that this replacement will here, systematically and unduly, favor irreversible decisions, for example, destroying the forests and building the highway. In the following representative case that we numerically explore, the size of this "irreversibility effect" appears fairly important.

I. Uncertainty and Improvement of the Available Information

Let T denote the number of periods to be considered in the model; period t starts at time $t-1$ and ends at time t. At time t, $t=0, \ldots, T-1$, two situations are possible: 1) A choice actually can be made between two decisions; then $v_t=0$ means that no irreversible decision is made at time t, whereas $v_t=1$ means that an irreversible decision is made at time t. 2) No choice actually remains open at time t; then we will write $v_t=1$. Hence we have: if t is such that $v_t=0$, then for every τ such that $\tau<t$, $v_\tau=0$; if t is such that $v_t=1$, then for every τ such that $\tau \geq t$, $v_\tau=1$.

Let us now express the fact that, as time passes, the decision maker gets more and better information about the state of the world, hence about the returns which constitute the consequences of his decisions. Let S_t denote the information held by the decision maker at time t; this information is valuable to him in so far as it concerns not only past and present returns—present meaning here "at time t"—but also future returns. At any time t different from the initial one, i.e., $t>0$, there can be several possible states $S_{t,e}$, $e=1, \ldots, N(t)$, of the information available to the decision maker at time t; the net discounted return at time t, i.e., from the t-th period, associated with $S_{t,e}$ will be denoted $R_{t,e}$. The information available at time t in state $S_{t,e}$ about the returns at times later than t is specified by means of transition probabilities from this state $S_{t,e}$ to the various possible states $S_{t+1,f}$, $f=1, \ldots, N(t+1)$, i.e., by means of conditional probabilities $p_{t+1,f}^{t,e}$ of the states $S_{t+1,f}$ given $S_{t,e}$; these transition probabilities may depend on the sequence $\{v_0, \ldots, v_t\}$ of decisions taken up to time t.

In the terms of Malinvaud, what we have defined is an "information structure" which simultaneously is "fixed" and "has memory." Malinvaud's information structure is defined in the following manner: he considers that at any time t, the states $S_{t,e}$ of the information available at t constitute a partition of the space—which can be identified with S_0—of the states of the world; to guarantee that his information structure has memory, he makes the assumption that the partition S_{t+1} is as least as fine as the partition S_t. Our information structure is more general than his in so far as he doesn't allow the partition S_t to depend on the sequence of decisions $\{v_0, \ldots, v_{t-1}\}$; moreover we are considering coverings and not only partitions.

I shall now consider a problem of sequential decision under irreversibility and uncertainty, where uncertainty is measured by the above information structure. To the initial random problem we will associate a riskless problem where, for every period and every sequence of decisions affecting this period, the random returns from this period given this sequence of decisions are replaced by their expected value calculated at time 0, i.e., at the time the decision maker solves his problem in order to decide what must be done immediately. We will see that a risk-neutral decision maker having solved the associated riskless problem more often adopts an immediate[5] irreversible decision

[4] Risk neutrality is unessential but simplifies matters; for a more general presentation allowing for any attitude towards risk, see my 1973 paper, which also examines choices with more than two alternatives.

[5] We are not interested in decisions which are not immediate, i.e., which would be made now and applied after a delay: indeed, after a delay another decision problem emerges, with another "time 0," requiring an immediate effective decision to which Proposition 1 eventually applies.

than does the same decision maker who has solved the initial random problem. This is an immediate consequence of Proposition 1.

PROPOSITION 1: *Consider a sequential decision problem under uncertainty, where irreversible decisions may be made and where the information structure is fixed and has memory: suppose the decision maker is risk neutral; consider the associated riskless problem. If the solution of the associated problem doesn't imply an immediate irreversible decision, then the solution of the initial problem doesn't either; but it may happen that the solution of the associated problem implies an immediate irreversible decision, whereas the solution of the initial problem does not.*

PROOF: In the initial problem, let $Q^*_{t,e}(v_0, \dots, v_{t-1})$ denote the maximum possible conditional expected value of future returns given the sequence $\{v_0, \dots, v_{t-1}\}$ of decisions; "conditional" refers to the state $S_{t,e}$ of the information available at time t and "future" means times t+1, ..., T. The principle of optimality[6] that must be followed by a decision maker solving the initial problem thus reads:

$$(1)\quad Q^*_{t,e}(v_0, \dots, v_{t-1}) = \max_{\{v_t \text{ compatible with } v_0, \dots, v_{t-1}\}} E_{t,v_0,\dots,v_t} \cdot [R_{t+1}(v_0, \dots, v_t) + Q^*_{t+1}(v_0, \dots, v_t)]$$
$$= \max_{\{\text{idem}\}} \sum_f p^{t,e}_{t+1,f}(v_0, \dots, v_t) \cdot (R_{t+1,f}(v_0, \dots, v_t) + Q^*_{t+1,f}(v_0, \dots, v_t))$$

In the associated riskless problem, let $\overline{Q}^*_t(v_0, \dots, v_{t-1})$ denote the maximum possible value of future (sure) returns given the sequence $\{v_0, \dots, v_{t-1}\}$; the principle of optimality that must be followed by a decision maker solving the associated riskless problem thus reads:

$$(2)\quad \overline{Q}^*_t(v_0, \dots, v_{t-1}) = \max_{\{\text{idem}\}} (\overline{R}_{t+1}(v_0, \dots, v_t) + \overline{Q}^*_{t+1}(v_0, \dots, v_t))$$

where

$$(3)\quad \overline{R}_{t+1}(v_0, \dots, v_t) = E_{v_0,\dots,v_t}[R_{t+1}(v_0, \dots, v_t)] = \sum_f p_{t+1,f}(v_0, \dots, v_t) \cdot R_{t+1,f}(v_0, \dots, v_t)$$

with

$$\forall t = 0, \dots, T-1,\ P_{t+1,f} = \sum_e P_{t,e} \cdot P^{t,e}_{t+1,f}$$

We will first show that $\forall t = 0, \dots, T-1$,

$$(4)\quad E_{v_0,\dots,v_{t-1}}[Q^*_t(v_0, \dots, v_{t-1})] \geq \overline{Q}^*_t(v_0, \dots, v_{t-1})$$

This result is trivial when $1 \in \{v_0, \dots, v_{t-1}\}$; furthermore equality then holds. Suppose now $1 \notin \{v_0, \dots, v_{t-1}\}$; then

$$(5)\quad Q^*_{t,e}(v_0, \dots, v_{t-1}) = \max_{v_t \in \{0,1\}} E_{t,v_0,\dots,v_t}[R_{t+1}(v_0, \dots, v_t) + Q^*_{t+1}(v_0, \dots, v_t)]$$

hence, $\forall v_t \in \{0, 1\}$,

$$(6)\quad E_{v_0,\dots,v_{t-1}}[Q^*_{t,e}(v_0, \dots, v_{t-1})] \geq \overline{R}_{t+1}(v_0, \dots, v_t) + E_{v_0,\dots,v_t}[Q^*_{t+1}(v_0, \dots, v_t)]$$

If $v_t = 1$, we now have

$$(7)\quad E_{v_0,\dots,v_t}[Q^*_{t+1}(v_0, \dots, v_t)] = \overline{Q}^*_{t+1}(v_0, \dots, v_t)$$

It is thus enough to show that

$$(8)\quad E_{v_0,\dots,v_{t-1},0}[Q^*_{t+1}(v_0, \dots, v_{t-1}, 0)] \geq \overline{Q}^*_{t+1}(v_0, \dots, v_{t-1}, 0)$$

As both members are zero for $t = T-1$, it is trivially true for $t = T-1$; it then results from

[6] About principles of optimality in stochastic dynamic programming see Arrow, note 4, pp. 524–25, and R. Bellman, pp. 199–200 and 205–09.

a recursive argument that it is also true for any $t \in \{1, \ldots, T-1\}$.

Hence, for $t=1$, we have

$$(9) \qquad E_{v_0=1}[Q_1^*(v_0 = 1)] = \overline{Q}_1^*(v_0 = 1)$$

$$(10) \qquad E_{v_0=0}[Q_1^*(v_0 = 0)] \geq \overline{Q}_1^*(v_0 = 0)$$

In the initial random problem, we choose $v_0=1$ if

$$(11) \qquad W_1 = E_{v_0=1}[R_1(v_0 = 1) + Q_1^*(v_0 = 1)] = \overline{R}_1(v_0 = 1) + E_{v_0=1}[Q_1^*(v_0 = 1)]$$

is greater than

$$(12) \qquad W_0 = E_{v_0=0}[R_1(v_0 = 0) + Q_1^*(v_0 = 0)] = \overline{R}_1(v_0 = 0) + E_{v_0=0}[Q_1^*(v_0 = 0)]$$

In the associated riskless problem, we choose $v_0=1$ if

$$(13) \qquad \overline{W}_1 = \overline{R}_1(v_0 = 1) + \overline{Q}_1^*(v_0 = 1)$$

is greater than

$$(14) \qquad \overline{W}_0 = \overline{R}_1(v_0 = 0) + \overline{Q}_1^*(v_0 = 0)$$

From (9), $W_1 = \overline{W}_1$; from (10), $W_0 \geq \overline{W}_0$.

For a case where $W_0 > W_1 = \overline{W}_1 > \overline{W}_0$, see the Appendix where it suffices to make:

$$(15) \qquad v_t = x_{t+1},\ x_{t+1} \in \{0, 1\}, \qquad t = 0, 1$$

$$(16) \qquad S_0 = \{\omega_e \mid e = 1, 2\},$$

$$\text{with } \begin{cases} y_1 = y_2 = -6 \text{ in the state } \omega_1 \\ y_1 = y_2 = +6 \text{ in the state } \omega_2 \end{cases}$$

$$(17) \qquad S_{1,e} = \{\omega_e\}, \qquad e = 1, 2$$

$$(18) \qquad S_{2,f} = \{\omega_f\}, \qquad f = 1, 2$$

$$(19) \qquad \overset{0}{p}_{1,e} = \tfrac{1}{2}, \qquad e = 1, 2$$

$$(20) \qquad \overset{1,e}{p}_{2,f} = \delta_{ef}, \qquad e = 1, 2, \quad f = 1, 2$$

$$(21) \qquad R_{t,1} = (x_t + 2)(4 - x_t - 6), \quad t = 1, 2$$

$$(22) \qquad R_{t,2} = (x_t + 2)(4 - x_t + 6), \quad t = 1, 2$$

$$(23) \qquad \overline{R}_t = (x_t + 2)(4 - x_t), \qquad t = 1, 2$$

Solving the initial random problem, the decision maker is led to adopt $v_0 = x_1 = 0$, whereas, by replacing this problem by the associated riskless problem, he is led to adopt $v_0 = x_1 = 1$.

It would be most interesting to have an idea of how large the irreversibility effect can be in a realistic case. With this aim in mind we have designed the following simulation. At time one $N(1) = 3^2$ different states of the available information are possible; they are denoted $S_{1,e(1)}$ where

$$(24) \qquad e(1) = (e_I(1), e_R(1))$$

$$(25) \qquad e_I(1) \in \{1 - \beta, 1, 1 + \beta\}$$

$$(26) \qquad e_R(1) \in \{1 - \gamma, 1, 1 + \gamma\}$$

furthermore for every $e(1)$ we have

$$(27) \qquad \overset{0}{p}_{1,e(1)} = 1/9$$

$$(28) \qquad R_{1,e(1)}(v_0 = 1) = \frac{e_I(1)}{1 + \sigma}$$

$$(29) \qquad R_{1,e(1)}(v_0 = 0) = F \cdot \frac{e_R(1)}{1 + \sigma}$$

This means that if an irreversible decision is made at time 0 (i.e., $v_0 = 1$) the undiscounted return at time one will be either $1-\beta$, or 1, or $1+\beta$, with probabilities 1/3, respectively; if no irreversible decision is made at time 0 (i.e., $v_0 = 0$), the undiscounted return at time one will be either $F \cdot (1-\gamma)$, or F, or $F \cdot (1+\gamma)$, with probabilities 1/3, respectively; the returns from $v_0 = 0$ are stochastically independent of the returns from $v_0 = 1$ (this is consistent with $N(1) = 3^2$).

At time two $N(2) = 6^2$ different states of the available information are possible; they are denoted $S_{2,e(2)}$ where

$$(30) \qquad e(2) = (e_I(2), e_R(2))$$

$$(31) \qquad e_I(2) \in \{(1-\beta)^a(1+\beta)^b \mid 0 \leq a+b \leq 2\}$$

$$(32) \qquad e_R(2) \in \{(1-\gamma)^c(1+\gamma)^d \mid 0 \leq c+d \leq 2\}$$

a, b, c, and d being nonnegative integers; furthermore for every $e(1)$ and every $e(2)$ we have (33)–(35). This means that, if for example $e_I(1) = 1-\beta$ and $e_R(1) = 1+\gamma$, there are only nine attainable states at time two, which are $(e_I(2), e_R(2))$ with

$$p^{1,e(1)}_{2,e(2)} \begin{cases} = 1/9 & \text{if } \begin{cases} e_I(2) \in \{e_I(1)\cdot(1-\beta),\ e_I(1),\ e_I(1)\cdot(1+\beta)\} \\ \text{and} \\ e_R(2) \in \{e_R(1)\cdot(1-\gamma),\ e_R(1),\ e_R(1)\cdot(1+\gamma)\} \end{cases} \\ = 0 & \text{otherwise} \end{cases} \tag{33}$$

$$R_{2,e(2)}(v_1 = 1) = \frac{e_I(2)}{(1+\sigma)^2} \tag{34}$$

$$R_{2,e(2)}(v_1 = 0) = F\cdot\frac{e_R(2)}{(1+\sigma)^2} \tag{35}$$

$$e_I(2) \in \{(1-\beta)\cdot(1-\beta),\ (1-\beta),\ (1-\beta)\cdot(1+\beta)\} \tag{36}$$

$$e_R(2) \in \{(1+\gamma)\cdot(1-\gamma),\ (1+\gamma),\ (1+\gamma)\cdot(1+\gamma)\} \tag{37}$$

If an irreversible decision has been made at time 0 or is made at time one (i.e., $v_1=1$), the undiscounted return at time two will be either $(1-\beta)\cdot(1+\beta)$, or $(1-\beta)$, or $(1-\beta)\cdot(1+\beta)$, with probabilities 1/3, respectively. If no irreversible decision is taken by time one (including, i.e., $v_1=0$), the undiscounted return at time two will be either $(1+\gamma)\cdot(1-\gamma)$, or $(1+\gamma)$, or $(1+\gamma)\cdot(1+\gamma)$, with probabilities 1/3, respectively. Again the respective consequences, of $v_1=1$ and of $v_1=0$ are stochastically independent.

At time t,

$$N(t) = \left(\frac{(t+1)\cdot(t+2)}{2}\right)^2$$

different states of the available information are possible; they are denoted $S_{t,e(t)}$ where

$$e(t) = (e_I(t),\ e_R(t)) \tag{38}$$

$$e_I(t) \in \{(1-\beta)^a(1+\beta)^b \mid 0 \le a+b \le t\} \tag{39}$$

$$e_R(t) \in \{(1-\gamma)^c(1+\gamma)^d \mid 0 \le c+d \le t\} \tag{40}$$

a, b, c, and d being nonnegative integers; furthermore for every $e(t-1)$ and every $e(t)$ we have (41)–(43).

For T=10 and different values of σ, β, γ, and F, we then have the following results: if $\sigma=0.05$ then $W_1=7.72$ and W_0 is given by Table 1. If $\sigma=0.10$ then $W_1=6.14$ and W_0 is given by Table 2.

As we have always chosen $\beta=\gamma$, the difference between W_0 and W_1 is a consequence only of: 1) the difference between 1 and F, and 2) the fact that W_0 results from a reversible decision, while W_1 results from an irreversible one.

Hence, for given σ, β, and γ (with $\beta=\gamma$), we may choose as a measure of the irreversibility effect the difference between 1 and that value of F, denoted $F_{IE}(\sigma, \beta, \gamma)$, which equalizes W_0 and W_1; for example

$$F_{IE}(0.10, 0.10, 0.10) = 0.87 \tag{44}$$

$$p^{t-1,e(t-1)}_{t,e(t)} \begin{cases} = 1/9 & \text{if } \begin{cases} e_I(t) \in \{e_I(t-1)\cdot(1-\beta),\ e_I(t-1),\ e_I(t-1)\cdot(1+\beta)\} \\ \text{and} \\ e_R(t) \in \{e_R(t-1)\cdot(1-\gamma),\ e_R(t-1),\ e_R(t-1\ \cdot(1+\gamma)\} \end{cases} \\ = 0 & \text{otherwise} \end{cases} \tag{41}$$

$$R_{t,e(t)}(v_{t-1} = 1) = \frac{e_I(t)}{(1+\sigma)^t} \tag{42}$$

$$R_{t,e(t)}(v_{t-1} = 0) = F\cdot\frac{e_R(t)}{(1+\sigma)^t} \tag{43}$$

TABLE 1

	$F=1$	$F=0.95$	$F=0.90$	$F=0.85$	$F=0.80$
$\beta=\gamma=0.05$	7.99	7.78	7.65	7.58	7.53
$\beta=\gamma=0.10$	8.26	8.03	7.84	7.69	7.57

TABLE 2

	$F=1$	$F=0.95$	$F=0.90$	$F=0.85$	$F=0.80$
$\beta=\gamma=0.05$	6.34	6.18	6.07	6.00	5.96
$\beta=\gamma=0.10$	6.54	6.36	6.20	6.08	5.99

as

$$(45)\quad W_0(\sigma=0.10, \beta=0.10, \gamma=0.10, F=0.87) = W_1(\sigma=0.10)=6.14$$

As might be expected, the irreversibility effect decreases with increasing σ (increasing σ meaning that future returns, hence uncertainty on future returns, are considered as less important) and increases with increasing $\beta=\gamma$ (i.e., with increasing uncertainty on future returns). As $\sigma=0.10$ is a usual choice for a discount rate and as $\beta=\gamma=0.10$ is a moderate estimation of uncertainty for many public investment programs,[7] 13 percent seems to give a good idea of the degree of magnitude of the irreversibility effect in a case where this effect can be considered as moderate.

APPENDIX

I shall use Theil's notations, assuming that T=2 (number of periods), $m=1$ (number of "instruments" or "controlled variables"), and $n=1$ (number of "noncontrolled variables" or "result variables"). The noncontrolled variable y_t, t=1, 2, which can be identified with Theil's "random element" $s(t)$, is subject to the following simultaneous probability distribution:

$$(A1)\quad Pr\,[y_1=y_2=-6]=Pr\,[y_1=y_2=6]=\tfrac{1}{2}$$

The "welfare function," i.e., the function whose expected value the decision maker wishes to maximize, reads

$$(A2)\quad \omega(x, y) = (x_1+2)(4-x_1+y_1) + (x_2+2)(4-x_2+y_2)$$

Irreversibility on the instrument x appears in the fact that the set where the decision maker will be allowed to choose a value for x_2 depends on the value he previously chooses for x_1, in the following way:

$$(A3)\quad x_1 \leq x_2$$

Let us solve this decision problem, taking into account that at the beginning of the second period, the decision maker will hold better information about the state of the world that obtains than he does at the beginning of the first period. Suppose he chooses $x_1 \in R$ at the beginning of the first period. If $y_1=y_2=-6$ obtains, he will then have to solve the following problem at the beginning of the second period:

$$(A4)\quad \max\,(x_2+2)(4-x_2-6)$$

under condition (A3), whose solution is

$$(A5)\quad \begin{cases} x_2 = -2 & \text{if } x_1 \leq -2 \\ x_2 = x_1 & \text{if } x_1 \geq -2 \end{cases}$$

If, on the other hand, $y_1=y_2=6$ obtains, the decision maker will then have to solve the following problem at the beginning of the second period:

$$(A6)\quad \max\,(x_2+2)\,(4-x_2+6)$$

under condition (A3), whose solution is

$$(A7)\quad \begin{cases} x_2 = 4 & \text{if } x_1 \leq 4 \\ x_2 = x_1 & \text{if } x_1 \geq 4 \end{cases}$$

Hence, to choose the appropriate value of x_1 at the beginning of the first period, the decision maker has to solve:

$$(A8)\quad \max G(x_1)$$

distinguishing three cases to explicit $G(x_1)$:

1) $x_1 \leq -2$; then

$$(A9)\quad G(x_1)=\tfrac{1}{2}[(x_1+2)(4-x_1-6)+0] + \tfrac{1}{2}[(x_1+2)(4-x_1+6) + (4+2)(4-4+6)]$$

[7] We consider cases where T=10; hence a single period in our scheme will in general correspond to more than one year; even more than ten years in the "highway versus forests" case, in which case $\beta=\gamma=0.10$ is not a moderate but a very low estimation of uncertainty.

hence, on $[-\infty, -2]$, $G(x_1)$ reaches a unique maximum $G(-2)=18$.

2) $-2 \leq x_1 \leq 4$; then

$$\text{(A10)} \quad \begin{aligned} G(x_1) = {} & \tfrac{1}{2}[(x_1+2)(4-x_1-6) \\ & + (x_1+2)(4-x_1-6)] \\ & + \tfrac{1}{2}[(x_1+2)(4-x_1+6) \\ & + (4+2)(4-4+6)] \end{aligned}$$

hence, on $[-2, 4]$, $G(x_1)$ reaches a unique maximum $G(0)=24$.

3) $4 \leq x_1$; then

$$\text{(A11)} \quad \begin{aligned} G(x_1) = {} & \tfrac{1}{2}[(x_1+2)(4-x_1-6) \\ & + (x_1+2)(4-x_1-6)] \\ & + \tfrac{1}{2}[(x_1+2)(4-x_1+6) \\ & + (x_1+2)(4-x_1+6)] \end{aligned}$$

hence, on $[4, +\infty]$, $G(x_1)$ reaches a unique maximum $G(4)=0$.

In order to maximize $G(x_1)$, the decision maker therefore chooses $x_1=0$ as the value of the instrument for the first period.

Let us now examine the associated riskless problem, i.e., the "certainty case" in Theil's words. The decision maker must now maximize

$$\text{(A12)} \quad (x_1+2)(4-x_1) + (x_2+2)(4-x_2)$$

under condition (A3), which clearly implies the decision $x_1=1$ for the first period.

REFERENCES

K. J. Arrow, "Statistics and Economic Policy," *Econometrica*, Oct. 1957, *25*, 523–31.

——— and A. C. Fisher, "Environmental Preservation, Uncertainty, and Irreversibility," *Quart. J. Econ.*, May 1974, *88*, 312–19.

P. W. Barkley and D. W. Seckler, *Economic Growth and Environmental Decay: The Solution Becomes the Problem*, New York 1972.

R. Bellman, *Adaptive Control Processes: A Guided Tour*, Princeton 1961.

G. Dreyfus, *Circulaire à Messieurs les Directeurs Départementaux de l'Equipement*, Direction des Routes et de la Circulation routière, Ministère de l'Equipement et du Logement, Paris, Jan. 20, 1970.

A. C. Fisher, J. V. Krutilla, and C. J. Cicchetti, "The Economics of Environmental Preservation: A Theoretical and Empirical Analysis," *Amer. Econ. Rev.*, Sept. 1972, *57*, 605–19.

C. Henry, "Option Values in the Economics of Irreplaceable Assets," Laboratoire d'Econométrie de l'Ecole Polytechnique, disc. pap., Paris 1973.

E. Malinvaud, "First Order Certainty Equivalence," *Econometrica*, Oct. 1969, *37*, 706–18.

H. A. Simon, "Dynamic Programming under Uncertainty with a Quadratic Criterion Function," *Econometrica*, Jan. 1956, *24*, 74–81.

H. Theil, "A Note on Certainty Equivalence in Dynamic Planning," *Econometrica*, Apr. 1957, *25*, 346–49.

Industrial Organization and International Trade: Some Theoretical Considerations

By LAWRENCE J. WHITE*

There has been an increased interest in the relationship between domestic industrial organization and international trade flows. For example, there have been some recent suggestions that the *U.S.* antitrust laws should be softened to permit mergers that might allow producers to compete more effectively in export markets and to withstand better the competition of imports by foreign producers.[1] The goal of these suggestions is to improve the current account of the American balance of payments. Perhaps the proper response to this type of suggestion is that flexible exchange rates should be the tool to correct balance-of-payments difficulties and that antitrust laws should be judged solely on their domestic merits. But the day of general acceptance of truly flexible exchange rates still seems to be far off, and in the present world of fixed or semi-fixed exchange rates, arguments that domestic policies—fiscal, monetary, and now, antitrust—should be modified to meet balance-of-payments goals will continue to be heard.

Unfortunately, there has been very little economics literature that has explored the relationship between domestic market structure and trade. It has been generally recognized that an open import regime—low tariffs and the absence of quotas—can improve domestic market performance, essentially by providing more competition for domestic producers.[2] Richard Caves and Ronald Jones (pp. 206–10) have also argued that the possibilities for exporting can reduce the domestic price charged by a monopolist, if he cannot discriminate between domestic and export markets. But no one has explored the possibilities of a reverse relationship in which the causality would run from domestic market structure to export- and import-competing performance.

This paper will demonstrate that there are theoretical reasons to expect a relationship that runs from domestic market structure to foreign trade flows. A firm with market power will face different incentives and behave differently with respect to these trade flows than would a group of competitors. Indeed, *under a number of circumstances the suggestion mentioned above that the U.S. antitrust laws be weakened may well be counterproductive: exports may well decrease and/or imports increase if a competitive industry is monopolized.*

The technique of this paper will be to compare the results yielded by a competitive market structure with those yielded by a monopoly. Section I will present the basic model. It will be assumed that the domestic and imported products are perfect substitutes and that the supply curve for the competitive industry is identical to the marginal cost curve for the monopolist; this last assumption is made so that exclusive focus can be placed on the effects of market power. Section II will offer some extensions and modifications of the basic model. In all cases, it will be assumed that the domestic producer or producers are only a small part of the world supply and cannot affect world prices. Hence, they take world prices as given and maximize accordingly.[3]

* Assistant professor of economics, Princeton University.

[1] See the *New York Times*, January 9, 1972, Section XII, p. 31; *New York Times*, Nov. 12, 1972, Section III, p. 7; and *Washington Post*, Apr. 14, 1973, p. E8. Senator Hart's Subcommittee on Antitrust and Monopoly decided to hold hearings on this issue in October 1973.

[2] For a recent theoretical statement and empirical test of this proposition, see Louis Esposito and Frances Esposito. See also Caves and Jones, pp. 206–10.

[3] To assume otherwise would introduce oligopoly interdependence problems, which are no more likely

I. The Basic Model

First, let us assume that the domestic market is wide open to the world economy, with no tariffs, transport costs, or other barriers to trade. In Figure 1, let SS be the industry supply curve and DD be the domestic demand curve.[4] Since by assumption the economy is wide open and the domestic producer or producers cannot affect the world price, the domestic monopolist becomes just one more competitor in the world marketplace. Hence, domestic market structure will not affect export or import performance. At any world price above P_c, say P_w, the country will be a net exporter and Q_1Q_2 will be exported, regardless of market structure. At any world price below P_c, say P'_w, the country will be a net importer and Q_3Q_4 will be imported, regardless of market structure.

Let us now close the domestic market with tariffs or other trade barriers, so that the domestic monopolist is free to set a domestic price that is higher than the world price. Let MR be the domestic marginal revenue curve. Facing P_w, the competitive industry's behavior would be unchanged: it would continue to produce OQ_2 with OQ_1 sold in the domestic market and Q_1Q_2 exported. But the monopolist will take advantage of the domestic demand curve and provide only Q'_m to the domestic market, at a price P'_m. He would provide Q'_mQ_2 to the export market.

Thus, if he can segment his markets, the monopolist will generate a higher level of exports than will the competitive industry (a difference of Q'_mQ_1). This comes, however, at the expense of a consumer's surplus loss of ABC and an income transfer of P'_mABP_w from consumers to the monopolist.[5] In effect, the restriction of supply to the domestic market releases extra low-cost supply that can profitably be sold in export markets.

But this situation of a high domestic price and a lower international price would be considered to be dumping.[6] Suppose that dumping is not permitted by international trading rules. Then the monopolist can no longer take advantage of his domestic market and simultaneously export. He faces a choice: he can either produce exclusively for the domestic market and forego exporting, or he can choose to export and also sell at home but forego the monopoly profits from the domestic market. In the latter case, he becomes just another competitor in the world market, and he exports the same amount as would a competitive industry. Which alternative he will choose will depend on the extra profits, above the world price, that could be made by selling exclusively in the domestic market, versus the extra producer's surplus to be made from selling in export markets, or rectangle P_mEFP_w versus triangle FGH. At best, then, he will export the same as a competitive industry; at worst, he will forego exporting entirely.

Now let us suppose that there are impediments to exports (foreign tariffs, transportation costs, etc.) by the domestic industry, so we are exclusively concerned with the likelihood of imports. Suppose that P_w represents the potential landed price of imports, inclusive of tariffs. If the industry is competitive, no imports will enter; Q_c will be produced and will be sold domestically at a price P_c, just as would happen in isolation. The monopolist, however, is prevented from exploiting his full monopoly position. The best he can do is to limit price at a level just below P_w and produce just above Q_1. If he is accurate, no imports will enter. Market structure

to admit a definitive solution for worldwide markets than for national markets.

[4] To the extent that the rising supply curve reflects rising factor prices, this implies monopsony power for the monopolist and would imply a worse comparative performance for the monopolist in the analysis which follows.

[5] In a disequilibrium situation, the welfare gain from the Q'_mQ_1 extra exports is presumably greater than the value of the extra exports themselves (otherwise, one would not especially care about export performance). This welfare gain would have to be weighed against the costs of the consumers' surplus loss and the income transfer. Also, the alternative costs of achieving the extra exports through export subsidies to the competitive industry should be considered.

[6] As Caves and Jones, pp. 212–15, rightly point out, dumping can arise from the situation described in Figure 1 with a rising marginal cost curve; dumping does not require a high fixed cost, low marginal cost situation.

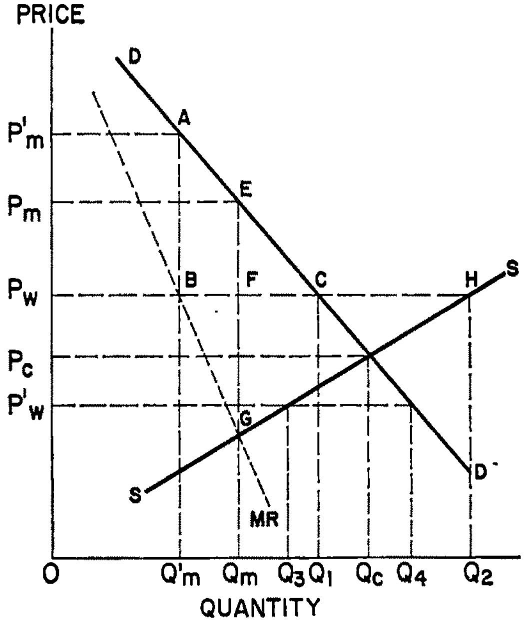


FIGURE 1

will not affect the level of imports. Finally, if the landed import price is P'_w, both the monopolist and the competitive industry will be limited to a price of P'_w and will produce Q_3; Q_3Q_4 in imports will enter. Again, market structure does not affect the level of imports.

These last results hold true, however, only if the foreign price and the domestic demand curve are known with certainty, so that the monopolist can limit price with certainty. Suppose, instead, that the import price is subject to uncertainty, due to uncertain exchange rates, uncertain foreign prices and costs, uncertain transportation costs, etc.[7] It will now be true that imports are more likely to enter under the monopoly regime than under the competitive regime. The monopolist's profit maximizing quantity will be lower, and hence will leave more room for imports, than is true for the competitive industry. The monopolist "lives more dangerously," with a smaller quantity, higher price, and more frequent imports, but he finds it more profitable to do things this way.

This proposition concerning the effects of uncertainty can be demonstrated as follows: Assume that the import price p_w has a probability distribution f and a cumulative probability distribution F. Assume that firms are risk neutral. Following Hayne Leland, let us assume that the domestic producer or producers are quantity setters; that is, at the beginning of each relevant planning period, the firm decides on and produces a specific quantity and then adjusts its price to meet the uncertainties of the period and clear the market.[8] That quantity, of course, implies a specific market-clearing price on the domestic demand curve p_d. Either of two results can follow: If the import price p_w turns out to be less than the implied domestic price, the import price is the ruling price, and imports will enter the country equal to the difference between domestic demand at the import price and the quantity that has been domestically produced. If the import price is equal to or greater than the implied domestic price, then the implied domestic price is in fact the ruling price and no imports enter. By assumption, the height of the import price *above* the implied domestic price does not affect the latter, since the domestic production has already been decided at the beginning of the period.

With a cost curve of $c(q)$, then, the monopolist profit function is

$$(1)\quad \pi = [p_d \cdot q - c(q)] \cdot [1 - F(p_d)] + \int_0^{p_d} f(p_w) \cdot [p_w \cdot q - c(q)] \cdot dp_w$$

In words, equation (1) says that once the monopolist decides on a quantity, his expected profits come from two alternative sources: the first term represents the profits

[7] Uncertainty has been introduced at this point for the first time because it would not have made any qualitative difference in any of the earlier results.

[8] The alternative would be to assume that the firms are price setters; that is, they set the price at the beginning of the period and adjust the quantity sold during the period. This latter alternative makes little sense in the face of uncertainty created by an uncertain import price. Logically, it would imply that the firm would sell zero units in the period in which it set a price that turned out to be higher than the import price.

that occur if the import price turns out to be above the implied domestic price and he is able to earn the implied domestic price (i.e., he has the domestic market to himself) times the likelihood that this will occur; the second term represents the profits (or losses) that occur when the import price is the ruling price, weighted by the likelihood of each import price.

Equation (1) can be rearranged to

$$(2) \quad \pi = p_d \cdot q \cdot [1 - F(p_d)] + \int_0^{p_d} f(p_w) \cdot p_w \cdot q \cdot dp_w - c(q)$$

Differentiating this with respect to the quantity of output that can be chosen by the monopolist, we get

$$(3) \quad \frac{d\pi}{dq} = \left(p_d + q \cdot \frac{dp_d}{dq}\right) \cdot [1 - F(p_d)] + p_d \cdot q \cdot [-f(p_d)] \cdot \frac{dp_d}{dq} + f(p_d) \cdot p_d \cdot q \cdot \frac{dp_d}{dq} + \int_0^{p_d} f(p_w) \cdot p_w \cdot dp_w - \frac{dc(q)}{dq}$$

Setting this equal to zero and rearranging terms, we find

$$(4) \quad MR_d \cdot [1 - F(p_d)] + \int_0^{p_d} f(p_w) \cdot p_w \cdot dp_w = MC$$

where MR_d is the marginal revenue of the domestic demand curve and MC is marginal cost. Equation (4) yields the not too surprising conclusion that the monopolist should equate marginal cost with marginal revenue, with the latter modified to a weighted average of the usual marginal revenue (when the import price is high enough to leave the domestic market solely to the monopolist) and the import price itself (when the import price is low enough so that the monopolist is forced to meet the import competition).

The corresponding rule for the competitive industry is achieved by setting equation (2) equal to zero. By simplifying and rearranging terms, we get

$$(5) \quad \hat{p}_d \cdot [1 - F(\hat{p}_d)] + \int_0^{\hat{p}_d} f(p_w) \cdot p_w \cdot dp_w = AC = MC$$

where $\hat{p}_d$ is now the implied market-clearing price for the competitive industry. The important difference between equations (4) and (5) is that the competitive industry equates price with marginal cost, even when the import price is above competitive levels. It is clear that the quantity that the monopolist chooses to satisfy (4) will be less than the competitive quantity that satisfies (5), as long as there is some positive probability that the imported price will sometimes exceed the implied domestic price at the quantity chosen by the competitive industry (i.e., $F(\hat{p}_d) < 1$).[9] But the lower quantity chosen by the monopolist will mean a larger expected level of imports. Thus, paradoxically, by choosing a smaller quantity and a higher implied domestic price, the monopolist thereby exposes himself more frequently to import competition—but he also makes some monopoly profits when the import price turns out to be high.

A similar result is achieved if we change the problem around somewhat and now assume that the import price is fixed and known with certainty but that the domestic demand curve is subject to uncertainty. Assume again that firms are quantity setters and are risk neutral. The basic mechanism at work is as follows: the domestic firm or firms set a quantity; either that quantity clears the market at a price below the import price, or demand proves strong enough so that the price rises to the ceiling

[9] For example, in Figure 1, suppose the average expected import price is P_w' with some (unshown) probability distribution around it. If this distribution does not extend as far up as the demand curve (i.e., $F(\hat{p}_d) = 1$), then both the competitive industry and the monopolist would choose output Q_3. But if the distribution of foreign prices extends above the demand curve (i.e., $F(\hat{p}_d) < 1$), then the competitive industry would choose an output somewhat below Q_3, and the monopolist would choose an output that was yet lower.

set by the import price and imports flow in to satisfy the remaining demand.

Let us assume that the domestic quantity q generates an expected price $\bar{p}_d$ on the domestic demand curve. The uncertainty of domestic demand is represented by an additional stochastic term p_d^*, which is added to or subtracted from $\bar{p}_d$. Assume that p_d^* has a probability distribution g and a cumulative probability distribution G, and that $g(p_d^*)$ is independent of $\bar{p}_d$ and q. Finally, the import price p_w is known with certainty.

The monopolist's profit function is

$$(6)\quad \pi=[p_w\cdot q - c(q)]\cdot[1 - G(p_w - \bar{p}_d)] + \int_{-\infty}^{(p_w-\bar{p}_d)} g(\overset{*}{p_d})\cdot[q\cdot(\bar{p}_d+\overset{*}{p_d}) - c(q)]\cdot d\overset{*}{p_d}$$

The monopolist's profits, once the quantity has been set, come from the instances when the import price is the ruling price, times the likelihood that this will occur, plus the instances when the domestic demand is weak enough so that his quantity alone clears the market, times the likelihood that this occurs.

Equation (6) can be rearranged to

$$(7)\quad \pi = p_w\cdot q\cdot[1 - G(p_w - \bar{p}_d)] + q\cdot\bar{p}_d\cdot G(p_w - \bar{p}_d) + q\cdot\int_{-\infty}^{(p_w-\bar{p}_d)} g(\overset{*}{p_d})\cdot\overset{*}{p_d}\cdot d\overset{*}{p_d} - c(q)$$

Differentiating with respect to q, we get

$$(8)\quad \frac{d\pi}{dq} = p_w\cdot[1 - G(p_w - \bar{p}_d)] + p_w\cdot q \cdot g(p_w - \bar{p}_d)\cdot\frac{d\bar{p}_d}{dq} + \left(\bar{p}_d + q\cdot\frac{d\bar{p}_d}{dq}\right)\cdot G(p_w - \bar{p}_d) - q\cdot\bar{p}_d\cdot q(p_w - \bar{p}_d)\cdot\frac{d\bar{p}_d}{dq} + \int_{-\infty}^{(p_w-\bar{p}_d)} g(\overset{*}{p_d})\cdot\overset{*}{p_d}\cdot d\overset{*}{p_d} - q\cdot g(p_w - \bar{p}_d)\cdot(p_w - \bar{p}_d)\cdot\frac{d\bar{p}_d}{dq} - \frac{dc(q)}{dq}$$

Setting this equal to zero and rearranging terms, we get

$$(9)\quad p_w\cdot[1 - G(p_w - \bar{p}_d)] + MR_d\cdot G(p_w - \bar{p}_d) + \int_{-\infty}^{(p_w-\bar{p}_d)} g(\overset{*}{p_d})\cdot\overset{*}{p_d}\cdot d\overset{*}{p_d} = MC$$

Again, appropriately defined marginal revenue equals marginal cost.

We get the corresponding rule for the competitive industry by setting (7) equal to zero and rearranging terms, or

$$(10)\quad p_w\cdot[1 - G(p_w - \hat{p}_d)] + \hat{p}_d\cdot G(p_w - \hat{p}_d) + \int_{-\infty}^{(p_w-\hat{p}_d)} g(\overset{*}{p_d})\cdot\overset{*}{p_d}\cdot d\overset{*}{p_d} = AC = MC$$

where $\hat{p}$ is the competitive price in this case. Again, the quantity that the monopolist chooses to satisfy equation (9) will be less than the competitive quantity that satisfies (10), but this means that more imports will enter under the monopoly regime.

This kind of model would be consistent with the kind of behavior that Laurence Krause found in the *U.S.* steel industry in the 1950's. Rising steel imports seemed to have no limiting effect on domestic steel prices. It may well be that this oligopolistic industry, facing uncertain import prices or uncertain domestic demand, chose to maximize profits in the ways indicated by equations (4) and (9), even though this meant high and rising import levels.[10]

II. Modifications and Extensions

One immediate extension of the basic model is to assume that the monopolist has lower costs than the competitive industry. Lower costs (greater efficiency) is the usual announced justification for mergers, and these may offset the effects of greater market power for the merged firms.[11]

[10] Less charitably, one might hypothesize that the *U.S.* steel industry simply miscalculated.

[11] See Oliver Williamson for a general cost-benefit approach for analyzing the market power versus efficiency gain tradeoff. It is worth emphasizing that in an equilibrium world, a dollar of exports or imports would be treated no differently from a dollar of domestic goods, and Williamson's approach would not require

An important distinction needs to be made between efficiency gains which reduce total costs but which do not affect marginal costs and those which reduce marginal costs. In the former case, the firm enjoys a windfall gain, but its short-run behavior at the margin is not altered. Many of the efficiency gains frequently mentioned in mergers—reductions in overhead costs, repair facilities, and inventory costs—appear to be of this nature. Hence, despite the efficiency gains, the analytics of Section I would still hold. But there may be some efficiency gains which do affect marginal behavior, or we may be interested in long-run effects, so it is worth exploring the case of lower costs for the monopolist.

It is readily apparent that, under conditions of certainty, the monopolist with lower costs will now generally outperform the competitive industry. In an open framework, the monopolist will export more or permit fewer imports. He is just one more competitor in a world of competitors, and his lower costs allow him to perform better. If discrimination is possible and dumping is permitted, the monopolist with lower costs will dump yet more exports into the world market. If dumping is not allowed, he may focus exclusively on the domestic market or he may export as in the open framework. The larger producer's surplus from the lower marginal costs makes the latter possibility more likely than was true when the monopolist had the same costs as the competitive industry.

If we reintroduce uncertainty and look at the import side, the picture is not quite so favorable to the monopolist. Expected imports will be reduced under the monopolist only if he is willing to produce a larger quantity than the competitive industry. This means that the implied domestic price for the monopolist must be less than that of the competitive industry. Suppose that the monopolist has an efficiency gain of X percent lower marginal costs compared to the competitive industry. Looking first at the case in which the import price is uncertain, we can determine the degree of efficiency gain that is required to produce favorable results for the monopolist. Letting e_d represent the (positive) elasticity of domestic demand, from (4) we get

$$(11)\quad p_d\left(1-\frac{1}{e_d}\right)\cdot[1-F(p_d)] + \int_0^{p_d} f(p_w)\cdot p_w\cdot dp_w = MC\cdot(1-X)$$

If p_d just equals the competitive industry's price $\hat{p}_d$, we can use equation (5) to simplify this to

$$(12)\qquad X=\frac{\hat{p}_d}{e_d}\cdot\frac{[1-F(\hat{p}_d)]}{MC}$$

Hence, if p_d is to be below the competitive price (and hence, if expected imports are to be smaller under the monopolist), the monopolist's efficiency gain of X percent reduction in marginal costs must be greater than the value of the right-hand side of equation (12). Thus, just the existence of an efficiency gain is not sufficient to reduce expected imports. It must be large enough to offset the increase in market power (represented by the elasticity term) that the monopolist gains, modified by the beginning probability that the import price will be above the domestic price and by the initial marginal cost.

For the case in which domestic demand is the uncertain variable, it can similarly be demonstrated that X must be greater than $(\hat{p}_d/e_d)\cdot G(p_w-\hat{p}_d)/MC$ for expected imports to be lower under the monopolist.

A second modification to the basic model is to assume that the domestic and foreign products are imperfect, rather than perfect, substitutes.[12] In this case, the monopolist

modification. But in a disequilibrium world, in which an improvement in the balance of payments becomes an important policy goal, the likely effect of a merger on trade flows may be worth including in the cost-benefit calculus.

[12] Besides differences in design or materials or the lack of consumer familiarity with the foreign product, other reasons for imperfect substitutability between the domestic and foreign products would be the vagaries of international shipping schedules, the possibilities of dock strikes and shipping strikes, the difficulties of suing

unambiguously will allow more imports to enter than would a competitive industry with identical costs. Since the demand for the domestic product is now less than infinitely elastic, the monopolist will set a higher price than will the competitive industry. But this will induce a greater demand for the imported product, hence more imports. On the export side, the monopolist's higher price will mean a lower volume but the effect on receipts will depend on elasticities of demand, and no a priori judgement can be made. Further, if the monopolist has lower costs, the general rule again holds that the reduction in costs must more than offset the increase in market power for the qualitative nature of these results to change.

Finally, we can ask what will happen if the domestic and foreign product are imperfect substitutes but the domestic industry has the ability to produce a product identical to the imported product. If the foreigner's capabilities and intentions are known with certainty, the model developed by Peter Swan indicates that imports will be equally limited by competitive or monopoly market structures—the former because of its low price on the initial domestic product and the latter because of its need to limit price and/or produce the new product to forestall entry by the foreigner.[13]

If the foreigner's intentions and capabilities are not known with certainty, however, this conclusion does not hold. The competitive result would be unaffected; the low competitive price would still forestall imports. But the monopolist would now have to weigh the possibilities that the foreign producer might not ship his goods after all. The monopolist would calculate the expected value of his own actions, given the expected probabilities of the foreigner's actions, and decide accordingly. Clearly there will be times when the monopolist will be unlucky and guess wrong: either he will cut his price on the first item and produce the new product and thus spoil his first market when in fact the foreigner was not going to enter the second market; or he will decide not to take these actions and then find that the foreigner has in fact entered the second market. This second case, though, means that imports will enter the domestic market. Once again, in the face of uncertainty, the monopolist has chosen to live dangerously, maximizing expected profits even though this means that imports will enter occasionally, whereas the competitive industry, with its competitive discipline and zero profits "rule," always has low prices and hence always keeps out imports.

This kind of uncertainty model appears to be relevant to the case of automobile imports. The *U.S.* automobile industry has been sluggish in providing small cars for the American market, hoping that *U.S.* consumers would therefore opt for larger, more profitable cars.[14] This might have worked if foreign producers had not been ready to ship small cars to the *U.S.* market. As it was, the *U.S.* manufacturers guessed wrong or miscalculated twice—in the late 1950's and in the late 1960's—and imports grew much faster than if small American cars had been available throughout this period.

III. Conclusion

The purpose of this paper has been to explore the connections between domestic market structure and international trade flows. We have demonstrated that under a number of interesting and reasonable circumstances, market structure can indeed influence trade flows. On the import side, a monopoly market structure is likely to allow *greater* import levels than would a competitive industry, unless the domestic and imported products are perfect substitutes and all variables are known with certainty. Even lower marginal costs by the monopolist may not help; the efficiency gain has to be great enough to overcome the effects of the market power that the monopolist can exploit. The export side is more ambiguous. A monopolist who can discriminate and dump may export more than a competitive industry. But if dumping is not allowed, he

under a different legal system in the event of breach of contract, etc.

[13] See also my 1972 paper.

[14] See my 1971 book, ch. 11.

might choose to focus his attention on the domestic industry and export less than the competitive industry. If the domestic and foreign products are imperfect substitutes so that the monopolist enjoys some market power even in export markets, then anything might happen, and one has to know the specific details of price elasticities and market positions before making any predictions.

Overall, changes in antitrust policy appear to be a very uncertain and inferior way of achieving improved performance in the international arena. In a number of likely circumstances they will be counterproductive. Antitrust policy should be based on its efficacy in improving consumer welfare in domestic markets. Exchange rate policy is indeed a far superior tool for dealing with balance-of-payments difficulties.

REFERENCES

R. E. Caves and R. W. Jones, *World Trade and Payments*, Boston 1973.

L. Esposito and F. F. Esposito, "Foreign Competition and Domestic Industry Profitability," *Rev. Econ. Statist.*, Nov. 1971, *53*, 343–53.

L. B. Krause, "Import Discipline: The Case of the United States Steel Industry," *J. Ind. Econ.*, Nov. 1962, *11*, 33–47.

H. E. Leland, "Theory of the Firm Facing Uncertain Demand," *Amer. Econ. Rev.*, June 1972, *62*, 278–91.

P. L. Swan, "Market Structure and Technological Progress: The Influence of Monopoly on Product Innovation," *Quart. J. Econ.*, Nov. 1970, *84*, 627–38.

L. J. White, *The Automobile Industry Since 1945*, Cambridge 1971.

———, "A Note on the Influence of Monopoly on Product Innovation," *Quart. J. Econ.*, May 1972, *86*, 342–45.

O. E. Williamson, "Economics as an Antitrust Defense: The Welfare Tradeoffs," *Amer. Econ. Rev.*, Mar. 1968, *58*, 18–36.

New York Times, Jan. 9, 1972; Nov. 12, 1972.

Washington Post, Apr. 14, 1973.

The Economics of Environmental Preservation: Comment

By Ronald Cummings and Virgil Norton*

In a recent paper in this *Review* by Anthony Fisher, John Krutilla, and Charles Cicchetti (F-K-C), an argument is developed concerning the preservation of natural environments which is fundamentally based on the assumption that in certain instances, conversion of natural environments for development purposes represents an *irreversible* investment. An economic model is proposed for the allocation of natural environments between preservation and development from which F-K-C conclude that ". . . it will in general be optimal to refrain from development even when indicated by a comparison of current benefits and costs if, in the relatively near future, . . . disinvestment, which is impossible, would be indicated" (p. 609). F-K-C continue with an interesting discussion of concepts regarding the measure of benefits and costs for evaluating projects involving the development of natural environments, and conclude with a case study of the Hells Canyon Project.

The purpose of this note is twofold. First, in Section I, we wish to suggest some serious weaknesses in F-K-C's rigid and sometimes ill-defined conception of irreversibilities, and to comment on the nature of the decisions which may be made if one accepts the direction suggested by F-K-C. In Section II, F-K-C's analytical model is extended to include a broader conception of the preservation-development problem. Concluding remarks are given in Section III.

I

We find a basic problem in the F-K-C paper which results from their nebulous and somewhat inconsistent distinction between preservation (P) and development (D) options, and the relationship of these to irreversible investments. F-K-C's examples of development include the following: the ". . . transformation and loss of whole environments as would result . . . from clear cutting a redwood forest, or developing a hydroelectric project in the Grand Canyon" (p. 605); the development of ". . . additional sites along the river, the construction of facilities to accommodate larger numbers of flat water recreation seekers, the penetration by roads of virgin sections, etc. . . . , an extinct species or ecological community that cannot be resurrected, a flooded canyon that cannot be replicated . . . " (p. 612).

In attempting to apply F-K-C's concept of development as implied by the diverse examples given above, consider a totally virgin area and the following sequence of possibilities. (a) Recreational use of the area is initiated but limited to hikers and backpacking with initial small investments made by the Forestry Service for clearing and marking specific hiking trails, and later, establishment of periodic shelters for camping. (b) The area is penetrated by access roads to allow more (low density) camping, but the "wilderness" nature of the area is maintained. (c) More roads are developed, higher intensity camping sites are provided, the area becomes a large recreation park. (d) The region's rivers are dammed for the purpose of power generation and flatwater recreation facilities are provided. Since F-K-C's D and P are, by assumption ". . . the highest valued use or combination of uses . . . " (p. 606), for D and P, it is necessary to distinguish between D (which according to F-K-C must be irreversible) and P; i.e., in the example above, which degree of investment is considered as D and which is considered as P for inclusion in the F-K-C model? Stated simply, in the sequence described above, when does P stop and D begin?

The implication of F-K-C's arguments is that D begins with irreversible investments;

* Professors of resource economics, University of Rhode Island. This article is contribution #1471 from the Rhode Island Agricultural Experiment Station.

they state, "Clearly, were the converse true, i.e., were the transformation [from P to D] reversible, much of the conflict between preservation and development would vanish" (p. 607).[1] Once again, however, in the example given above, where does irreversibility begin? F-K-C argue, somewhat curiously, that "... irreversibility of development is fundamental to the problem" (p. 612), but need not be absolute. Two kinds of reversibilities are possible: the restoration of an area by a program of direct investment (but this has little relevance for the sorts of phenomena of concern to them) and a natural reversion to the wild, which they view as of little relevance to their main concerns.

Thus, the definitions of P and D as well as the "economics of preservation" as seemingly viewed by F-K-C, encompass those problems associated with *technically*[2] irreversible investments which affect natural environments. We view as important a recognition of this limitation of the F-K-C approach and would like to submit the following comments for consideration.

We grant that it may be technically impossible to restore a wooded area containing camp sites or a flooded canyon to their *exact* original state.[3] It is not clear, however, that such exactness is a prerequisite for the future generation of recreational benefits. Even a flooded canyon may be restored to *some* kind of a "natural" environment at a later date, and the issue is one of benefit-losses (or possibly gains?) associated with two forms of an open-space, or wilderness, environment.

Therefore, with the exception of an extinct species, we find it difficult to conceive of irreversible investment options,[4] but further, we are not convinced that it is useful to do so. It seems to us that the assumption, if not imposition, of irreversibility vis-à-vis the use of natural environments abstracts from a whole set of issues of paramount interest to the economist.

F-K-C's arguments concerning increasing future demand for recreation due to rising incomes may be extended to argue that future generations will be capable of paying higher costs for *reversing* earlier development projects. Should we not therefore be inquiring as to alternative project designs that might be developed which allow tradeoffs between current efficiency (present benefits) and less costly future reversals? It is important to note that the approach taken by F-K-C, i.e., that of assuming the highest valued use for each P and D, forces the choice to a specific D. This prevents the possibility of choosing a lower valued but relatively reversible use of D.

Finally, a host of issues are relevant concerning equity and income distribution. For example, in some cases it may be argued that major users of wilderness areas are relatively wealthy and have access to various alternative areas and types of recreation; the "development" of a given wilderness area *could* result in income (taxes or charges) which could be used elsewhere for the establishment of parks for ghetto children.[5]

II

We have argued that technical irreversibilities of investments for development must be viewed as a most special case, but that costs of reversals may be so large relative to benefits as to make investments for development *economically* irreversible. The latter statement requires empirical measure, however, and suggests the need for an analytical

[1] This distinction becomes fuzzy later in the paper, however. F-K-C later view D as including simply an increase in the intensity of use of a recreational area (see p. 612). This inclusion is somewhat odd given the arguments which follow.

[2] "... a program of direct investment [for reversing D] ... would seem to have little relevance ... for the sorts of phenomena with which we are mainly concerned ..." (p. 612).

[3] William Lord suggests that "... every decision is irreversible in the strict sense that all previous conditions can never be restored exactly. By the same token, it is difficult to conceive of a situation in which some of the previous conditions could not be restored by some conceivable alternative action" (February 9, 1973 letter to the authors).

[4] Nor, unless one insists on the *exact* replication of an area by investment, do the examples given in F-K-C make the task simpler.

[5] It is tempting to draw an analogy between F-K-C's irreversibility arguments relating to natural resources and to human resources; e.g., are environmental effects (which may be affected by transfers) on ghetto children irreversible?

model for the generation of such measures. Following F-K-C's suggestion, p. 612, we wish to offer such a model, and demonstrate that the F-K-C optimization approach to the preservation-development allocation may be viewed as a special case of an optimization problem for the intertemporal determination of production and investment rates for natural resource industries in general. Both ends may be realized by simply applying the production-investment model for natural resource industries given in Oscar Burt and Cummings to the environment-preservation problem.[6] To facilitate comparisons, F-K-C's notation and assumptions are used. We wish to maximize:

$$(1)\quad \sum_{t=1}^{T} [B_1^t(P^t) + B_2^t(D^t) - I^t - G^t]\beta^t$$

subject to the restrictions

$$(2)\qquad D^{t+1} = D^t + \sigma I^t - \gamma G^t$$

$$(3)\qquad P^{t+1} = P^t + \gamma G^t - \sigma I^t$$

$$(4)\qquad P^t + D^t = L$$

all $0 \leq t \leq T$, all variables are nonnegative.

In (1) through (4), I^t is investment in development, as in the F-K-C paper; G^t is investment in preservation, i.e., G^t converts developed land into a natural, or preserved, environment, and is a convenient method for allowing reversibilities for I^t;[7] B_1 and B_2 are benefit functions for preservation and development, respectively; β^t is the discount factor, $(1+r)^{-t}$. Equations (1), (2), and (4) correspond to F-K-C's equations (1), (4), and (2), respectively. For completeness, we include an explicit transition equation for preserved land P^t, recognizing that the interdependence between P and D would allow its elimination.

Assume that at least some small portion of the natural environment L is in D and P during all periods, i.e., D^t and P^t are positive for all t.

Using (1) through (4), the following Lagrangian expression is formed.

$$(5)\quad H = \sum_{t=1}^{T} \Big\{ [B^t(P^t) + B^t(D^t) - I^t - G^t]\beta^t - \lambda^{t+1}\beta^{t+1}[D^{t+1} - D^t - \sigma I^t + \gamma G^t] - \Gamma^{t+1}\beta^{t+1}[P^{t+1} - P^t - \gamma G^t + \sigma I^t] - \alpha^t\beta^t[P^t + D^t - L] \Big\}$$

Maximization of H requires conditions which include the following:[8]

$$(6)\qquad \Gamma^{t+1} = \sum_{\tau=t+1}^{T} \left(\frac{\partial B^\tau}{\partial P^\tau} - \alpha^\tau\right)\beta^{\tau-(t+1)} + \Gamma^{T+1}\beta^{T-t}$$

$$(7)\qquad \lambda^{t+1} = \sum_{\tau=t+1}^{T} \left(\frac{\partial B^\tau}{\partial D^\tau} - \alpha^\tau\right)\beta^{\tau-(t+1)} + \lambda^{T+1}\beta^{T-t}$$

$$(8)\quad (1+r) = \sigma(\lambda^{t+1} - \Gamma^{t+1}), \qquad \text{if } I^t > 0$$
$$(1+r) > \sigma(\lambda^{t+1} - \Gamma^{t+1}) \rightarrow I^t = 0$$

$$(9)\quad (1+r) = \gamma(\Gamma^{t+1} - \lambda^{t+1}), \qquad \text{if } G^t > 0$$
$$(1+r) > \gamma(\Gamma^{t+1} - \lambda^{t+1}) \rightarrow G^t = 0$$

Conditions (8) and (9) imply that $(I^t)(G^t) = 0$; i.e., at any t, investment for development *and* preservation may not take place. Equation (8) may be used to deduce F-K-C's development in their equations (17) through (20); equation (8) is, of course, their equilibrium condition (14) with $I^t > 0$.

In equations (6) and (7), Γ^{T+1} and λ^{T+1} may be shown to measure the marginal value of terminal stocks of preserved and developed land,[9] and may be treated as zero,

[6] In the interest of conserving space, the entire Burt-Cummings model is not repeated here.

[7] As suggested to the authors by Darrell Hueth, we have earlier argued that investment G^t which reverts land from D to P may not be exact; i.e., earlier environments may not be exactly replicated. In such cases a unit problem arises in terms of G and P which may be corrected either by introducing a factor f which converts G into units of P ($P^{t+1} = P^t + \gamma f G^t - I^t$ in equation (3)) or by using two state variables for P: P_1 for the natural environment and P_2 for the "near" or "man-made" environment. Further, as Charles Howe has suggested to the authors, the taxonomy of P and D precludes forms of I^t which simply involve more intensive utilization of the same acreage. This form of I^t, however, may remove land from P via externalities as opposed to actual land occupancy.

[8] See George Hadley, pp. 190–93, or Burt and Cummings, pp. 579–82.

[9] See equations (8) and (9) in Burt and Cummings, p. 580.

particularly as T becomes large.[10] The coefficients Γ^{t+1} and λ^{t+1} measure, respectively, the present value (evaluated at t) of the flow of benefits in all future periods associated with an incremental change in preserved and developed land in period t. Their difference, of course, $\lambda^{t+1}-\Gamma^{t+1}$ or $\Gamma^{t+1}-\lambda^{t+1}$ measures the *net* return from an increment of developed or preserved land in t, respectively.

The economic interpretation of the decision rules (8) and (9) is immediately obvious. If *net* benefits to development are at least as great as the marginal cost of development (which implies $G^t=0$), development is carried to the point where marginal costs and benefits are equated; using F-K-C's terminology, I^t lies in a free interval and G^t *must* lie in a blocked interval. Similarly for the *reversal* of development, G^t, $G^t>0$ implies that I^t lies in a blocked interval. It is plausible to expect that during some periods *net* returns to I or G may be positive, but less than the marginal costs of investment. During such periods I^t and G^t are both zero—they both lie in F-K-C's blocked intervals.

The model given in (1) through (4) may serve several purposes. First it allows for explicit consideration of future costs of reversibility—a consideration which we have argued deserves major attention. Second, some insight is provided for the problem which was not clear to F-K-C: the potential flexibility which reversibility may give to current investment policy. Finally, and perhaps most important, this view of the preservation-development investment problem provides for the input expected from economists, viz., the evaluation of a wide range of alternatives in time, as opposed to the "all or nothing" decisions which may result from the F-K-C irreversibility framework.

[10] The problem of allowing $T\rightarrow\infty$ is discussed in the Appendix to Burt and Cummings.

III

In conclusion, we argue that F-K-C's interesting arguments regarding environmental preservation not only apply to an extremely limited set of circumstances, which exclude the bulk of examples suggested in their paper, but also have the potential of encouraging decision makers to overlook flexible or reversible investments. If (as we suggest the reader of F-K-C has every reason to conclude) F-K-C's major concern is with technically irreversible investments, we suggest that these problems arise only with an extremely rigid definition of reversibility—specifically, one that insists on an exact resurrection of a natural environment. We are not convinced that social benefits from recreation are materially affected by such exactness, in which case the environmental preservation argument concerns *economic reversibility.* Viewing the problem in this broader perspective opens, we suggest, a number of lines of inquiry which may be useful in future evaluations of development projects.

REFERENCES

O. R. Burt and R. Cummings, "Production and Investment in Natural Resource Industries," *Amer. Econ. Rev.*, Sept. 1970, *60*, 576–90.

A. C. Fisher, J. V. Krutilla, and C. J. Cicchetti, "The Economics of Environmental Preservation: A Theoretical and Empirical Analysis," *Amer. Econ. Rev.*, Sept. 1972, *62*, 605–19.

G. Hadley, *Nonlinear and Dynamic Programming*, Reading 1964.

The Economics of Environmental Preservation: Comment

By A. EUGENE ABRASSART AND DALE D. MCFARLANE*

The recent article in this *Review* by Anthony Fisher, John Krutilla, and Charles Cicchetti (F-K-C) is an important step towards the development of a general method of analysis of the economics of environmental preservation. They are also commended for their attempt to adapt their model to the investigation of an important environmental policy question concerning the advisability of hydroelectric development in the Hells Canyon region of the Snake River. We feel, however, that there is room for improvement in their theoretical models and perhaps in the presentation of empirical results.

Sections I and II of their article consist of the development of an allocation model of some degree of sophistication from which some general conclusions are drawn which are not necessarily true. For instance, in Section I, it is stated that " . ... the marginal opportunity costs of development, the benefits from preservation, are increasing as development increases" (p. 607). This certainly follows from the assumptions upon which their model is based; i.e., from their assumptions it follows that $B^P_{PD}>0$. It is not completely relevant, however, for as their later analysis indicates, the important question is what happens to total benefits rather than marginal. With respect to the portion of Hells Canyon in question, the present level of developed area is very small. Since almost all of the area is in a preserved state, i.e., $P\cong L$, their assumptions $B^D_D>0$ and $B^D_{DD}<0$ imply that the marginal benefits of development are relatively large. Thus in seeking an optimal allocation of the land between preservation and development, we would seek to increase marginal benefits from preservation and decrease marginal benefits from development. Thus we would seek to increase the developed area D. We might also point out that an undeveloped Hells Canyon may be so unique a natural resource that, as F-K-C suggest, there are no adequate substitutes of like quality and as such their procedures are justified. In general, however, we should note that theirs is basically a suboptimization procedure in that it isolates its analysis upon a single area without regard to other areas. Thus, if such a method were to be used in a piecemeal area by area decision process, the end result may be far from optimal for the total environment.

In Section II, F-K-C use the concept of total benefits and conclude that "... as benefits from preservation increase relative to benefits from development, the optimal short-run level of development $D^*(t)$ decreases" (p. 611). The justification for this statement is provided in their footnote 17. There they correctly derive the relation

$$B^D_D = B^P_P e^{\alpha' t} \tag{1}$$

At this point they argue that "As t increases, $e^{\alpha' t}$ increases, so that B^P_P (the marginal benefits of preservation) must be decreasing, implying that P^* is increasing—and D^* decreasing" (p. 611).

The error in this conclusion may be seen if we recognize that both B^D and B^P are functions of time, specifically $B^D(D(t), t)$ and $B^P(P(t), t)$. Hence B^D_D and B^P_P are functions of time, so that if we wish to determine how B^P_P changes as t increases, we need only take the partial derivative of equation (1) with respect to t, as follows:

$$\begin{aligned} B^D_{Dt} &= B^P_P(\alpha' e^{\alpha' t}) + e^{\alpha' t} B^P_{Pt} \\ &= e^{\alpha' t}[\alpha' B^P_P + B^P_{Pt}] \end{aligned} \tag{2}$$

* Associate professors of management science, respectively, Oregon State University. Copies of Federal Power Company (*FPC*) exhibits and testimony were obtained while the authors acted as consultants to legal counsel for the power company applicants immediately prior to, and during, the *FPC* hearings.

From this it follows that

$$B^P_{Pt} = B^D_{Dt}e^{-\alpha' t} - \alpha' B^P_P$$

Now B^P_P decreases as t increases if, and only if, $B^P_{Pt}<0$; i.e., if, and only if,

$$B^D_{Dt}e^{-\alpha' t} - \alpha' B^P_P < 0 \tag{3}$$

or

$$B^D_{Dt} < \alpha' e^{\alpha' t} B^P_P \tag{4}$$

or, by (1)

$$B^D_{Dt} < \alpha' B^D_D \tag{5}$$

Thus B^P_P will decrease as t increases *only* if inequality (5) holds true. Although it was indicated by F-K-C that "... both theoretical and empirical considerations suggest that benefits from development are likely to be decreasing ..." (p. 609) over time, this merely indicates that $B^D_t<0$ and does not mean that $B^D_{Dt}<0$. Thus there is no theoretical or empirical reason to expect $B^D_{Dt}<\alpha' B^D_D$, which means there is no reason to expect $D^*(t)$ to be monotone decreasing. We should also note that since $B^D_D>0$ and presumably $\alpha'>0$, then for a given value of B^D_D, the larger the value of α' the more likely that the inequality $B^D_{Dt}<\alpha' B^D_D$ will be satisfied. As we shall show later, the assumption of a linear demand curve for wilderness recreation will tend to overestimate the growth rate α'. A more realistic assumption, namely that of curvilinear demand curve, would lead to a smaller value of α' and a smaller likelihood that the required inequality will be satisfied.

The possibly incorrect assumption that $D^*(t)$ is monotone decreasing is then used by F-K-C to determine the effect on optimal policy. Since we cannot be sure this assumption is correct, we cannot be sure that the policy that follows, namely that "the projected development in the Hells Canyon ... should be undertaken immediately, if at all" (p. 611) is, in fact, optimal.

Furthermore, even if it should be the case that inequality (5) is satisfied, the F-K-C policy, *ceteris paribus*, does not say that development should be undertaken immediately or not at all. "Immediately" in this case is merely time t=0 which might be any datum point in time. Thus, although there is presently a moratorium upon hydroelectric development in Hells Canyon, it may be true five or ten or more years hence that development should be undertaken immediately.

Section III of their article is devoted to a comparison of estimated benefits for alternative uses of Hells Canyon or some portion thereof. While the benefits of development are relatively easy to identify, the benefits of preservation are not. Therefore they take the apparently reasonable approach of asking what the preservation benefits would have to be to equal the benefits of development. They have already shown in Section II that preservation benefits can be expected to increase over time. They now show how development benefits can be expected to decrease over time. This is due to the effects of technological change as demonstrated in the model provided in the Appendix of their article. The technological change model is an improvement over the traditional analysis which is also detailed in their Appendix. They claim that "... gross hydroelectric benefits will be overstated between 5 and 11 percent when technological change is not introduced ..." (p. 613). Taken together, the general idea of growing preservation benefits and diminishing development benefits provide a powerful economic argument in favor of prolonged preservation. Unfortunately the figures obtained in the application of these models to the Hells Canyon area do not necessarily lead to the F-K-C conclusions that "... the area is likely to yield greater benefits if left in its natural state" (p. 617).

Consider first the idea that hydro benefits will be overstated by 5 to 11 percent in the traditional analysis. These figures are obtained from Table 1, Exhibit R-670, Federal Power Commission (*FPC*) hearings, and are said to result from the model given in the Appendix to the F-K-C paper. In this model, K is the traditional measure of hydro bene-

fits and $PVC_{1\ldots T}$ is the measure which incorporates technological improvement. The relationship between these is said to be

$$(6)\quad \frac{K}{PVC_{1\ldots T}} = \sum_{t=1}^{T} \frac{b_0}{(1+i)^t} \div \sum_{t=1}^{T} \frac{b_0/(1+\pi)^t}{(1+i)^t}$$

It is then claimed that[1]

$$(7)\quad \frac{K}{PVC_{1\ldots T}} = \frac{T}{\sum_{t=1}^{T} \frac{1}{(1+\pi)^t}}$$

Clearly if equation (6) is correct then equation (7) must be incorrect, i.e., not an equation. The error seems to have resulted from mistakenly assuming the sum of products to equal the products of sums. A proper simplification of $K/PVC_1 \ldots T$ that follows if equation (6) is correct is

$$(8)\quad \frac{K}{PVC_{1\ldots T}} = \frac{i(1+\pi)+\pi}{i} \cdot \frac{1-(1+i)^{-T}}{1-(1+i)^{-T}(1+\pi)^{-T}}$$

We mention this solely as a warning to the reader that one cannot expect to obtain correct values from the use of equation (7). Nor for that matter, can one expect to obtain the F-K-C data through the use of either equations (6) or (8) above or equation (27) of F-K-C. This is because π, the simplified representation of technological change for the development alternative, appears in each of these equations. Thus while these formulations will undoubtedly be of use in certain circumstances, they will not reproduce the F-K-C data. Their data were obtained by a more complicated process said to be outlined in their Appendix. It appears, however, that the results shown in their Table 1 were obtained in a manner somewhat different than as indicated. The traditional measure of the benefits of a hydro power project described on page 609 of F-K-C are said to be adjusted for technological progress. However the adjustment outlined in the Appendix is a year-to-year adjustment. This suggests that a different development benefit for each year is to be discounted back to present value to obtain the present value of development (adjusted) of Table 1. Exhibit R-671 of the *FPC* hearings reveals, however, that these figures were actually obtained by discounting a constant annual benefit (say, \$1,304,000) to present value for each of the first five years, a different constant annual benefit (say, \$1,647,000) to present value for each of the next ten years, and a third such annual benefit (say, \$644,000) to present value for each of the remaining thirty-five years to obtain $PVC_{1\ldots T}$ (say, \$13,809,000). One should also note that the $PVC_{1\ldots T}$ of their Table 1 is not the same thing as $PVC_{1\ldots T}$ mentioned in their Appendix, as the former includes other factors not included in the latter. We suspect these variations and the figures that result were appropriate in the particular project being studied. Nonetheless it is unfortunate that the data do not follow from the model as presented.

If we accept that their calculations of $PVC_{1\ldots T}$ are approximately correct as indicated above, there remain steps in their analysis which indicate that they may in fact have underestimated the preservation benefits necessary to forego development. This amount, b_p^m, is found by dividing the present value of development (adjusted) by the factor

$$\sum_{t=1}^{T} \frac{(1+\alpha_t)^t}{(1+i)^t}$$

Their term α_t represents a composite growth rate in preservation benefits. It is the sum of two other growth rates. This is dependent upon the assumption of a linear demand curve which undergoes both a vertical shift and a horizontal shift over time. The resulting growth rate α_t is appreciably larger than would be the case if the demand curve were curvilinear. Thus the rate of increase of

[1] A recent discussion with Krutilla verified that the equation as presented in F-K-C Appendix is the result of a typographical error and the intended form is as shown in our equation (7).

benefits may be substantially overestimated and therefore b_p^m substantially underestimated.

In addition to the underestimation of necessary preservation benefits, there is reason to suspect that an overestimation of future preservation benefits is inherent in their analysis. As was mentioned earlier, an overestimation of the growth rate of preservation benefits results from their assumption of a linear demand curve. Basically the model proposed by F-K-C is based on the concept of consumer surplus with a shifting demand schedule as demand increases through time.

Linearity is a convenient assumption in economic analysis, but there are times when this assumption results in gross discrepancy in empirical studies. It is the authors' contention that a better approximation of the as-yet-undetermined demand curve would have been obtained through the assumption of a non-linear demand curve. Such demand curves should be familiar to those interested in attempts to estimate demand schedules for recreational activities. Even references cited by F-K-C relating to empirical studies support the nonlinearity contention.[2] Differences in the slope of two curves, one linear and the other non-linear, at the appropriate intercept values result in substantial differences in area when the demand schedule is shifted through time. Both a parabolic and a exponential curve with the degree of curvature held constant can be shifted through the appropriate intercept values in a manner approximating F-K-C analysis. However, instead of the 15 percent rate of growth in benefits indicated by F-K-C, our preliminary analysis shows that a rate of growth approximately one-half that great would be achieved using a demand schedule with a degree of curvature consistent with previously cited empirical studies.

The magnitude of error that results from assuming a 15 percent rate of growth as opposed to say a 7 percent rate can be immense when the compounding effect is considered over a long time span. For example, after the first twenty years, the ratio of total benefits received under the linear assumption (15 percent rate of growth) to total benefits received under the non-linear assumption (7 percent rate of growth) would be in the neighborhood of 2.5 to 1, and after fifty years, the ratio would be 17 to 1. It follows that the initial year's benefit necessary to warrant preservation will be much larger than those given (say, \$80,122) when such benefits are calculated using non-linear demand schedule.

More important from a theory point of view is their use of the concept of consumer surplus as measured by the area under the demand curve. It is proper to use such a measure under conditions of constancy in the marginal rate of substitution of income for wilderness recreation for any given quantity of the latter. This will occur in the event that the marginal utilities of income and wilderness recreation are independent and the marginal utility of income is constant, or alternatively, that both marginal utilities may change but must change in the same direction and proportion. The F-K-C analysis fails to provide a justification for making either assumption. On the contrary, their assumption that preferences will change over time suggests that the marginal utility of money will change from period to period. Unless appropriate changes in the marginal utility of wilderness recreation also occur, the marginal rate of substitution will vary from one indifference curve to the next for a given quantity of wilderness recreation. Even this possibility seems dispelled by their assumption that, presumably apart from preference changes, growth in per capita income results in increased quantities of wilderness recreation demanded. That is, F-K-C implicitly regard wilderness recreation to be a normal good. That being the case, it follows that the use of the area under the demand curve will result in an overstatement of consumer surplus.[3] This, of course,

[2] See William Brown, Ajmer Singh, and Emery Castle.

[3] See Kenneth Boulding. Also see Michael E. Burns for a discussion of path-dependency difficulties that result from simultaneous income and price changes as found in F-K-C demand curve shifts. E. J. Mishan also

will be true whether the demand curve is linear or non-linear. Thus, the F-K-C analysis overestimates the benefits of preservation. This, in turn, causes their estimate of necessary preservation benefits to be further understated.

Finally, in comparing necessary preservation benefits obtained from their model with estimates of actual benefits, it should be noted that both the *Review* article and the Krutilla testimony at the *FPC* hearings are devoid of data representing benefits of preservation beyond the initial period, t=0. Data for later periods are shown in Table 1 below.

TABLE 1—GROWTH OF PRESERVATION BENEFIT

Time	Visitor Days/ Year	Total Benefits/ Year	Benefit/ Visitor Day
t=0	121,000	$ 895,00	$ 7.40
t=20	814,000	15,976,000	19.63
t=50	814,000	118,370,000	145.42

These have been obtained using the F-K-C assumptions, parameter values $k=20$, $r=.04$ $i=.09$, $r_y=.05$, $\gamma=.10$, and what F-K-C claim is a conservative estimate of initial preservation of $895,000 per year. Even the relatively modest figure of $80,122, cited as the required initial year's benefit, implies later years' benefits that are substantial—in excess of $10 million in year fifty. We feel it would be better not to base a policy decision solely on the required initial year's benefit, as this ignores the question as to whether or not the hypothesized growth rates and the required later benefits that follow from them are reasonable.

In summary, while we recognize the complexity of analysis associated with the economics of environmental preservation and admire the F-K-C attempt to provide a framework for analyzing such problems, we find their article to be a perplexing amalgamation of insight and contradiction. The mathematics used range from sophisticated to incomplete or incorrect. The economic theory is innovative in its application but marred by oversights and conclusions inconsistent with the premises. The empirical work is a welcome testing of the theory in an area important both for the Hells Canyon in particular and the environment in general. Unfortunately, much of the data necessary to verify their results were not provided. In addition many of the results did not follow from the model(s) presented.

We would be remiss if we failed to make the final point that we recognize that there are many variables which of necessity are excluded from the analysis; for example, the benefits received from those of us who may never visit Hells Canyon but whose personal utility preferences are that the region be preserved in its natural state. Such preservation might well be justified, exclusive of the results of the F-K-C analysis. Our point is merely that their analysis in its present state should not be used for policy decision making.

shows how the money index of consumer surplus may increase as real income increases even though actual utility remains unchanged.

REFERENCES

K. E. Boulding, "The Concept of Consumer Surplus," *Amer. Econ. Rev.*, Dec. 1945, *35*, 851–69.

W. G. Brown, A. Singh, and E. N. Castle, "Net Economic Value of the Oregon Salmon-Steelhead Sport Fishery," *J. Wildlife Manage.*, Apr. 1965, *29*, 66–79.

M. E. Burns, "A Note on the Concept and Measure of Consumer's Surplus," *Amer. Econ. Rev.*, June 1973, *63*, 335–44.

A. C. Fisher, J. V. Krutilla, and C. J. Cicchetti, "The Economics of Environmental Preservation: A Theoretical and Empirical Analysis," *Amer. Econ. Rev.*, Sept. 1972, *62*, 605–19.

E. J. Mishan, "Interpretation of the Benefits of Private Transport," *J. Transp. Econ. Policy*, May 1967, *1*, 184–89.

Federal Power Commission (*FPC*), "In the Matter of: Pacific Northwest Power Company and Washington Public Power Supply System," Hearings, Washington 1970.

The Economics of Environmental Preservation: Further Discussion

By ANTHONY C. FISHER, JOHN V. KRUTILLA, AND CHARLES J. CICCHETTI*

The comment by Ronald Cummings and Virgil Norton (C-N) on our (F-K-C) paper focuses on the role of irreversibility in resource development projects in natural environments. C-N pose the questions: Is irreversibility a technical or an economic concept, and if the latter, is not the F-K-C approach overly restrictive? In other words, even if it is true that it is not possible to restore a flooded canyon to its original state (which C-N concede), would it not be possible to invest in the production there of a "natural" environment of comparable value, should this prove desirable? On the assumption that this is indeed the case, C-N go on to perform an analytical exercise spelling out its implications for investment in both development and "preservation" in natural environments. In principle, we would agree that this sort of exercise may be a legitimate extension of what C-N call the "special case" treated by us. There are, however, a number of problems with their formulation, primarily in their perception of the problem in its empirical setting. In the brief space permitted, we shall try to bring out some of these problems.[1]

I

Taking a wilderness area, C-N postulate a sequence of investments for recreational improvements such as shelters, access roads, high density recreation facilities (including a dam), and so on, and ask where in the sequence P stops and D begins. P, of course, stops before the C-N improvements begin; that is, under the Wilderness Act of 1964, none of the investments are permitted, nor are motorized vehicles, whether road or off-road vehicles. The legislation seeks to establish preserves to make provision for a diversity of outdoor recreation preferences. These areas are intended to meet the demands of that submarket that seeks in an outdoor recreation experience solitude in an undisturbed natural environment, remoteness and privacy in a wilderness camp setting, etc. The question that C-N appear to be asking is what is the optimal mix of fresh air and carbon monoxide in a wilderness setting. Admittedly the imputation is motivated by our feeling that C-N are victims of a fallacy of composition. The problem that needs to be addressed is how to allocate resource-based outdoor recreation assets so that those with a low carbon monoxide tolerance as well as those who respond negatively to opportunities for pedestrian heroics each have an opportunity to do his thing.

Perhaps some perspective on the allocation problem can be obtained by noting the relative distribution of resource-based outdoor recreation assets. The federal land management agencies currently administer roughly 20 percent of the area of the coterminous United States. Somewhere between 2 and 3 percent of the total falls within the *roadless category*, a criterion for possible inclusion in the national wilderness system. Considering, then, a given roadless tract of land, subject to evaluation under the Wilderness Act, or possibly under the Wild and Scenic Rivers Act, the question is not what is the optimal degree or disruption of its natural environment, but rather which of a number of allocation alternatives are to be preferred. For example, should the area be: (a) preserved in its undisturbed natural condition, (b) reserved for some nonwilderness type recreational activities that may be compatible also with some selective resource extraction or development, or (c) devoted to the extraction of resource commodities with-

* University of Maryland, Resources for the Future, Inc., and University of Wisconsin, respectively. We are grateful to Talbot Page and Kerry Smith for a number of helpful suggestions.

[1] For a fuller discussion, see Fisher and Krutilla.

out providing for recreation? One would look at the prospective returns associated with each of the alternatives in order to evaluate the opportunity costs of any mutually exclusive management options (see the F-K-C empirical analysis). But one should be aware that there are some 350 million acres of roaded National Forest areas, National Parks, and other Interior Department lands (equal in area to the combined Common Market Countries, excluding the United Kingdom) that can readily accommodate the kind of recreation C-N visualize. Perhaps one should not dismiss casually the thought that the opportunity cost of preserving a large part of the 30 to 50 million acres of roadless areas and wild and scenic rivers in an undisturbed condition may not be very large in comparison with the precluded wilderness recreation and other values attendant on converting de facto wilderness of diminishing supply to the large and growing stock of roaded and otherwise developed resource-based recreational lands.

The issue then turns on economic reversibility, as described in our opening paragraph. This is an empirical issue, of course, but it seems to us that C-N are prone to base their judgments on too simplistic a view of the vast and complicated outdoor recreation market, and indeed overlook what empirical evidence we have about this and related markets. Outdoor recreationists are not an undifferentiated mass seeking a homogeneous commodity. The implied carburetor policy, i.e., the appropriate mix of gas and air, is likely to satisfy only a fraction of the total market—just as the "on foot and fresh air" policy will satisfy only a part of the multimodal distribution of tastes in outdoor recreation. The significant work of George Stankey indicates that there is even a bimodal distribution in tastes, forming a purist-nonpurist dichotomy within the ranks of *wilderness* recreationists. This touches on the C-N hypothesis that exact replications of natural environments, wilderness faunal associations, etc., are unimportant to recreation seekers. It may be true for the bulk of the market for outdoor recreation that will have some 350 million other acres on which to indulge such interests, but it is not true of the important and rapidly growing "purist" submarket.

The matter of authenticity as an attribute in the demand for undisturbed natural environments may be likened to the demand for authenticity in the visual arts. For the bulk of the art museum clientele, the difference between an original work of art by one of the masters and an exact forgery by one of his protégés, or a contemporary art student, is undetectable, and may well satisfy the bulk of the population. But for a connoisseur, the mere suspicion of a forgery so expertly done that even art critics will differ in their opinion as to its authenticity will be significant and will result in a drastic reduction in the market value of the objet d'art, as many museum curators have been embarrassed to discover.

The reason for the unsatisfying nature of even relatively accurate forgeries for the purist is not, of course, a matter of economic knowledge, and perhaps not even of scientific knowledge generally. There may be an aura about the works of creative genius that the work of the most gifted imitator cannot provide. There may be even a cult comprised of those who revere the works of nature in a sense similar to that in which the Buddhists revere it, which may not be dissimilar from the reverence that many primitive societies confer on nature in their religious observances. To those who number among the purists, preservation of the constituents of the biosphere in precisely the way in which it has evolved, without disturbances from post-industrial man, is a matter of great significance in a profound personal sense. Such feelings, indeed, have been captured in the works of Wordsworth, Emerson, and Thoreau in very moving fashion. Whatever the reasons may be, whether mystical or religious, they are felt with great intensity.[2] For ana-

[2] Anyone having attended a public hearing on proposed wilderness designations, proposed extractive activities in de facto wilderness, etc., where representatives of the Sierra Club, say, or the Wilderness Society have testified will have obtained a flavor of the intensity of such feelings.

lytical purposes this maps into a highly inelastic demand for the "originals." This clientele group represents a significant market that appears to be growing rapidly as income, educational, and urban composition of American society changes (see Stankey and Cicchetti). Moreover, this is a market for which refinements in restorative technology will do little by way of recreating "undisturbed" natural environments. Accordingly, the argument for irreversibility, whether technical or economic, is a powerful one where the clientele group place a high value on the attribute of authenticity in the amenity services yielded by natural environmental resources.

II

We have reflected on environmental authenticity in recreation in response to C-N. In fact recreation constitutes only one of the uses of undisturbed natural environments, as has been discussed elsewhere (see Krutilla (1967a)). Considering another of these uses, namely the provision of a laboratory and materials for research in the life and earth sciences with often valuable application in the fields of medicine and agriculture, we may arrive at a better understanding of irreversibility in the context of environmental resource use. Thus in response to the quotation from William Lord in C-N, although it is obviously true that investment in specialized plant and equipment, for example, represents an irreversible commitment of capital, this is of no greater importance to society than the irreversibility which attends the death of any individual member of a species. Loss of the scientific and technological information needed to produce an alternative sort of specialized equipment, for the services of which there may be future demand, is however a matter of more consequence, and may be likened to loss of the genetic information contained within the last viable mating pair of the species. Extinction of a species, or more generally destruction of a unique resource, represents a possibly important reduction in the options available to society, and illustrates a central postulate of welfare economics: expansion of choice represents a welfare gain; reduction of options, a loss.

III

Even if survival of a species is not at issue, it may be misleading to represent, in the fashion of C-N, investment in "preservation," in an environment like the Hells Canyon of the F-K-C empirical study, a virgin redwood forest, or some other with attributes not readily producible. The clear cutting of a climax species such as the redwoods, for example, is equivalent to the removal of the results of ecological succession over a period of centuries. A climax species removed would be succeeded by various seral species in a procession of changing plant and animal communities, culminating in the original ecological relationships only after a lapse of time much greater than the ordinary economic period of production. This is recognized in the formal analysis of F-K-C through the introduction of the concept of rate of reversion, represented by the symbol δ, of a transformed environment to its natural state.

It is not, however, recognized in the C-N model in which it is implicitly assumed that the mere expenditure of funds can produce at any moment any amount of preservation desired. This can be seen in their equations (2) and (3), where γ represents the number of units of "preserved" environment that can be purchased for a unit of investment expenditure. Not only is there no known technology, human or otherwise, to recreate in any conventional period of production a Hells Canyon, but on the basis of the evidence concerning the preferences of users of this type of resource there cannot be even in principle, any more than there can be a recreation of an "authentic" Rembrandt. C-N view the problem as involving a continuum of reversible decisions and criticize F-K-C for casting it in discontinuous form. This misses the point that a natural environment, like a work of art, can't be cut into small pieces, used, and then restored to its original condition, with value unimpaired.

This is not simply a problem of an excessively long period of investment. Once it

is recognized that the transformation of Hells Canyon, the deepest gorge on the North American continent with its unique biotic associations, as contemplated in the project to build one or more large dams and related facilities there, cannot be reversed via investment on a human scale to restore anything like the original environment—any more than could the clear cutting of a virgin redwood forest, or other such transformation—the C-N model runs into further difficulty. It is simply not true that investment expenditures G in their equations (2) and (3) map into increments to P, the measure of undeveloped area—either with respect to the attributes of the area, or the preferences of users. Nor is the difficulty overcome by the suggestion in footnote 7 of a mechanical insertion into equation (3) of an additional factor f which "converts G into units of P." The problem is precisely our inability to achieve this transformation either technically or economically in important cases such as the one we studied, and no amount of additional factors attached to investment expenditure G in an equation describing the change in the stock of land in state P is likely to make it go away.

C-N's other suggestion in footnote 7, to specify two state variables, P_1, the "natural environment," and P_2, the "man-made environment," is not necessarily wrong in principle, and in our judgment deserves further study. It should be noted, however, that it implies that the conversion from P_1, once made, cannot be reversed, just as in our article. Moreover, it would introduce substantial additional complexity into the model, without any indication that this would be repaid by any gain in insight or applicability. That is, since P_2 and P_1 are mutually exclusive alternatives, why not simply classify the P_2 uses as one of the non-preservation options, to be balanced against preservation, as in F-K-C? In fact in the Hells Canyon analysis the developed recreation benefits were considered along with power benefits. However, since the incremental costs for recreation were greater than the incremental benefits this component was eliminated in our final calculations.

Finally, let us address the applicability or usefulness of the two models. We might have anticipated an objection, especially from applied researchers or resource managers, that the informational requirements for application of our model are so great that it may not be operational in a real situation. Instead, we have in the comment by C-N a suggestion which, if taken seriously, would multiply by many times (as many as the number of types and degrees of preservations, $P_2, \ldots P_n \ldots$) the informational requirements of the model for project planners. Again taking the Hells Canyon project and our analysis as an example, what additional empirical analysis, based on what data, would C-N undertake along the lines of their suggestion?

This latter point may be viewed another way. C-N suggest that we have erred by not taking into account the costs and benefits of reversing development. Assuming for a moment that somehow investment in preservation is feasible, if costs of reversal are incurred some number of years in the future (say fifty years), the present value of these costs is negligible with any reasonable rate of discount. Yet the present value of preservation benefits occurring within the first fifty years in the Hells Canyon analysis represented the bulk of the total. Reversing development, as C-N suggest, would add little to either the present value of the costs, or the present value of preservation benefits. Accordingly, it would not alter the basic conclusions of the F-K-C empirical analysis. Of course, this would not be true for a reversal very early in the economic life of the project, but this would clearly be wasteful, wiping out most of the benefits after virtually all of the costs have been incurred.

IV

It appears that in economic policy these days cutting a tree is as likely to save a ghetto child as anything else that appears likely to be done. But tomorrow will bring another day and it may be that we shall be able to prescribe better targeted programs of aid for the underprivileged (Milton Friedman and others have suggested giving them

money) than merely cutting trees—or destroying Hells Canyon. Accordingly, we defer to another day our comments on the equity issue as posed by C-N.

There is another equity issue however. Many people believe that certain environmental assets are jointly owned by all generations. This raises the question, does the present have the right to destroy such assets for its own benefit? Taking into account intertemporal ownership would lead to more preservation than assuming the present owns everything.

V

Although we still believe that our distinctions between reversible and irreversible resource uses are conceptually meaningful and useful to resource managers as working approximations, we acknowledge that the issue raised by C-N is a legitimate one for professional concern. The same cannot be said for the matters treated in the comment by A. Eugene Abrassart and Dale McFarlane (A-M). As we do not wish to permit the impression that there is any validity in their assertions of error in either our theoretical or empirical results, we find it necessary in our response to consider and refute each of these assertions. For the reader who may be (understandably) impatient with this procedure, it can be stated that, in brief, none of the results in the original analysis are affected by anything in A-M.

We feel obliged however to be especially careful and thorough because more is at stake than pride of authorship. In fact, the future of Hells Canyon itself is at stake. Our previous analysis has been a part of the Federal Power Commission (*FPC*) proceedings and our present work along with work done by A-M may well become part of the decision process.

Let us now turn to the specifics, which we address as briefly as possible and more or less in order of their appearance in A-M. To begin with, A-M suggest that optimally, we should seek to increase development in Hells Canyon on the basis of their presumption that the marginal benefits from development there are "relatively large," as "almost all of the area is in a preserved state." In the first place, "relatively" here is with respect to marginal benefits when development is already substantial. Diminishing marginal returns to development ($B^D_{DD}<0$) says not a thing about a comparison with marginal benefits from preservation, which, in any real situation, could exceed the returns to development at any level of development—or vice versa.

In addition to this logical error, A-M are also wrong on empirical grounds. As it happens, the upper reaches of Hells Canyon are already developed by three Idaho Power Company dams, and the lower reaches would be virtually completely developed by either of the alternatives proposed by the Pacific Northwest Power Company.

They next assert that we isolate our analysis "upon a single area without regard to others." In fact, all benefit and cost estimates for the alternative uses of the Hells Canyon depend explicitly upon the availability and substitutability of resources from other areas, as a review of the original study (pp. 609–15) will indicate.

There follows a fairly long and confused discussion of our demonstration (p. 611) of the implications for the optimal short-run level of development D^* of the projected alternative use benefit streams (pp. 609–10) and the concavity assumptions (p. 607). The point of this part of our analysis was embarrassingly simple, and can be stated quite generally: the lower the marginal *net* returns to an activity, the lower the optimal level (of the activity). For example, if the (assumed negatively sloped) marginal revenue curve of a firm shifts down and the (assumed positively sloped) marginal cost curve shifts up, the firm's optimal output will be reduced. What we did (fn. 17, p. 611) was simply to show that the corresponding curves (marginal benefits and marginal opportunity cost of development in Hells Canyon, respectively) must behave in this fashion over time, again, given the benefit projections and concavity assumptions also accepted here by A-M.

The source of their confusion may be some sloppy notation in the original. Let us re-

write equation (a) of footnote 17 as

$$(1) \qquad B_t = B_0^P(P)e^{\alpha t} + B_0^D(D)e^{-\beta t}$$

where B_t is the flow of benefits at time t, B_0^P and B_0^D are benefits at time $t=0$ as functions of P and D, respectively, growing at rates α and $-\beta$, respectively. Optimal instantaneous development D^* is found by differentiating this expression with respect to D and setting equal to zero:

$$(2) \quad \frac{dB}{dD} = \frac{dB_0^P(P)}{dD} e^{\alpha t} + \frac{dB_0^D(D)}{dD} e^{-\beta t} = 0$$

Substituting $(L-D)$ for P in the first term on the right-hand side in order to carry out the differentiation, substituting back again, and rearranging, we obtain

$$(3) \qquad \frac{dB_0^D(D)}{dD} = \frac{dB_0^P(P)}{dP} e^{(\alpha+\beta)t}$$

which, letting $(\alpha+\beta)=\alpha'$, is just equation (b) of footnote 17. Clearly, as t increases $e^{\alpha' t}$ increases, and it follows that D^* must be decreased in order to increase the marginal benefits from development (left-hand side) to balance the increase in $e^{\alpha' t}$ (right-hand side). A-M assert that dB_0^D/dD and dB_0^P/dP are functions of time, and proceed accordingly to take partial derivatives with respect to t. This is wrong, however. B_0^D and B_0^P are simply the benefit functions (of D and P, respectively) defined at time $t=0$. The dependence of B_t on time is not felt in $B_0^D(D)$ or $B_0^P(P)$, but only in the exponential $e^{\alpha' t}$ which expresses the relative rate at which the P-benefit flow is growing on its base at $t=0$, appropriately indicated in equations (1)–(3) above by the subscript. Granted some lack of precision in the original notation (it may be that A-M were misled by the notation of equation (a)), the point of the discussion was, we feel, intuitively obvious. In any case, let us emphasize that the implications of the projected behavior of benefits, along with diminishing returns, are exactly as stated earlier: namely, optimal instantaneous development D^* must be decreasing over time.

Concluding their discussion of the theoretical portion of our study, A-M note that "... the F-K-C policy, *ceteris paribus*, does not say that development should be undertaken immediately or not at all." This is not what we said. What we said (p. 617) was that an optimal (present) decision would be to proceed with the development immediately, if at all. We also indicated quite explicitly that "although a particular program, in this case nondevelopment, may be indicated given current anticipations, it can be revised (in the direction of further development) at any time following the emergence of new and unanticipated relationships in the economy, as for example, a reversal of the historic decline in [real] energy costs" (pp. 611–12).

VI

The A-M assertion that the data obtained from application of our models to Hells Canyon do not necessarily lead to our conclusion that "the area is likely to yield greater benefits if left in its natural state" (p. 617) is incorrect. The data lead unequivocally to this conclusion. What A-M attempt to do is to require that an explicitly simplified representation of our technical change adjustment computational model, used solely for heuristic purposes and notational compactness in the text (equation (27)), be used directly in reproducing the data in Table 1 by inserting the numerical rate of technical change for π (exchange of correspondence). That is, of course, absurd. It is nonetheless true as A-M claim, that there is a problem with the mechanics of the simplification in the Appendix. The last equation there, as reproduced in equation (6) of A-M, is in error. The summation before the denominator should not have been inserted; instead the entire quotient should have been enclosed in brackets with the summation applicable to the quotient rather than separately for both the numerator and denominator. But, conceding that, the carelessness reflected in developing the compact notation is without any effect on either the theoretical or the empirical results presented in the original study.

Next, while A-M concede (under prompting) that a more complex method was used in developing data in Table 1 than by use of equation (27) directly, they argue that the data were developed in some way inconsistent with our technical change adjustment computational model given in the Appendix. This also is incorrect. What they have confused is the need to accommodate the special conditions for power benefit evaluation in the Columbia River system, of which the hydroelectric site in Hells Canyon is a part, with inconsistency between the model and actual benefit estimation. In the Pacific Northwest, the relation between hydro potential and regional power demands has been, until now, such as to permit virtually all of the region's power needs to be met from hydroelectricity. As regional power demands grow to exceed the hydroelectric potential, however, the system will change from an exclusively hydro to a mixed hydro-thermal and eventually to a predominantly thermal system with a hydro peaking component. In each phase of the region's power development the hydro system will be operated differently; initially to maximize prime power production, secondly to provide "dependable capacity" (the same volume of water passed through greater turbine capacity, for a shorter period of time higher in the load curve to emphasize capacity, rather than energy production), and finally to peaking capacity predominantly, while the thermal component carries the base of the load (see Krutilla (1967b)). Since the hydro energy and capacity benefits will differ for each phase of operation, three different energy-capacity values were *stipulated* in the Federal Power Commission (*FPC*) proceedings, based on the three power system simulations appropriate to the three phases provided by the Bonneville Power Administration. Accordingly three different representative annual benefit values were used for obtaining the present values of Table 1, reflecting the changes over time in the role of the hydro component, and hence the proposed hydroelectric project, in the region's power system. But each of the representative values was adjusted by 5–11 percent depending on the set of assumptions used for each of the various cases, *precisely as determined by the technical change adjustment computational model given in the Appendix to our paper.* We wish to emphasize, therefore, that what is reflected in the process of estimating the power benefits is simply common sense in adjusting the application of the general model to the special circumstances of a particular case. And we specifically reject the A-M assertion that "Many of the results did not follow from the model(s) presented."

VII

We turn now to the vague suggestion that the assumption of linear demand schedules for recreation in the computational model is inconsistent with the findings of one of our references, William Brown, Ajmer Singh, and Emery Castle, and more seriously, that it results in overstatement of the rate of growth of recreation benefits, and therefore the size of the benefit stream over time. Taking the first point, Brown et al. estimated both linear and non-linear (exponential) demands. The fit was approximately the same in each case, though Brown et al. exhibited some small preference for the exponential, on the grounds that the linear resulted in an *under*estimate of benefits (p. 277). Beyond this, there is no theoretical argument advanced in support of nonlinearity anywhere in the literature. On the contrary, the only studies we are aware of in which the specification of functional form of recreation demand equations is influenced by theoretical considerations argue for linearity.[3]

We nevertheless remain open-minded about whether the exponential form cited by A-M and estimated (along with the linear) by Brown et al. may be somehow more appropriate than the linear, but observe that it does not in any case imply the difference in benefit growth rates claimed by A-M. Consider the exponential equation

$$y = ae^{-bx} \tag{4}$$

[3] See Oscar Burt and Durward Brewer, and Cicchetti, Fisher, and V. Kerry Smith.

where y is quantity demanded, x is price, and a and b are positive constants. The intercepts with y and x axes are $(0, a)$ and $(\infty, 0)$, respectively. More precisely, the curve doesn't actually intersect the x-axis, rather

$$\lim_{x\to\infty} ae^{-bx} = 0$$

The area under the curve between $x=0$ and $x=\infty$ is given by

$$(5) \qquad \int_0^\infty ae^{-bx}dx = \lim_{n\to\infty} -\frac{a}{b}(e^{-bx})\Big|_0^n = \frac{a}{b}$$

Now suppose the shift parameter a is increased k-fold, where k is any constant. Then (4) becomes

$$(4') \qquad y = kae^{-bx}$$

the new intercepts are $(0, ka)$ and $(\infty, 0)$, and the area under the curve is

$$(5') \qquad \int_0^\infty kae^{-bx}dx = \lim_{n\to\infty} -\frac{ka}{b}e^{-bx}\Big|_0^n = \frac{ka}{b}$$

We observe that, exactly as in the linear case, the sum of the percentage rates of growth in the intercepts is equal to the rate of growth in area. The y-intercept is increased from a to ka, an increase of $(k-1)$ 100 percent, the x-intercept remains at ∞, and the sum of the percentage rates is

$$(k-1)(100) + 0 = (k-1)(100) \text{ percent}$$

which is just equal to the rate of growth in area.[4]

To conclude discussion of the linear versus exponential issue, we note a serious error in A-M's jump from rate of growth to size of benefit stream. That is, even were we to accept their contentions, (a) that a non-linear demand is preferable to a linear on either theoretical or empirical grounds, and (b) that use of the non-linear will result in a smaller benefit growth rate, we would not obtain anything like the linear-non-linear benefit ratios they present, and for a very simple reason. They have assumed indefinite compounding (at the asserted different rates) of benefits, whereas we have explicitly set out (pp. 613–16) a series of projections involving a dampening of the growth rates, and the reasons for this.

Before passing on to the next issue, use of the consumers' surplus concept, we might speculate briefly as to the source of the discrepancy between A-M's assertions and the results reported in equations (4)–(5′) and footnote 4. It may be that they have assumed a non-linear curve passing through the same intercepts, initially, as a linear. Such a (convex) curve would of course cut off a smaller area, initially, than a linear, and indefinite undamped growth on this smaller base would produce a declining non-linear-linear benefit ratio over time. Our computational method was however to normalize on an initial benefit area of \$1, which of course implies initial intercepts varying with the slope and shape of the demand curve. In particular, both (initial) intercepts of a convex curve such as the exponential or parabolic would have to be larger than those of a linear in order to cut off the same area. The question is, then, what is the relationship between the rates of *growth* in these intercepts and the rate of *growth* in area and this we have addressed above.

VIII

Now on the matter of consumer's surplus, A-M have fallen into an elementary textbook error: confusion of movement along a demand curve with a shift in the curve. Consumer's surplus, measured along an income-compensated demand curve, provides a money measure of the utility change associated with an own price change. For this measure to be uniquely determined in any one period, it is sufficient only that either (a) demand is income-compensated, or (b) all income elasticities of demand are unitary, or (c) a small fraction of the budget is spent

[4] It is not clear that this exact relationship holds when a curve such as $y=ae^{-bx}-1$ is chosen to produce a finite x-intercept which shifts with the curve. In our calculations, which are rather tedious, it is either exactly or approximately satisfied, depending on the choice of parameters. Similar results are obtained for the other non-linear form mentioned by A-M (though not estimated by Brown et al.), the parabolic.

on the good(s) in question (see for example, Cicchetti et al., Harold Hotelling, and Eugene Silberberg). Of course, even where none of these conditions hold, the area under the demand measure provides a useful approximation to the money evaluation of the utility change, as convincingly demonstrated by Michael Burns.

Over time, technical change can (as we have assumed) result in growth in income, which in turn increases the quantity demanded at every price. This shift in demand is not, however, to be confused with movement along a demand schedule having any of the properties described above, or for that matter not having them.

If none of conditions (a), (b), or (c) holds, then it is true that the area under the static demand curve for a normal good does not coincide exactly with the Hicksian consumer's surplus. But though it is greater than the compensating variation, it is smaller than the equivalent variation, which we have suggested (p. 616) may be the appropriate measure in this case. It is their confusion on this point which leads A-M to present us with what might be characterized as the "baloney" theory of benefit measurement, which they gratuitously attribute to Kenneth Boulding, Burns, and E. J. Mishan. According to this theory, only goods for which the income elasticity of demand is zero, or perhaps negative (as in baloney) are legitimate subjects for applied welfare analysis.

IX

Finally, a comment on our culpability for not providing data for years yet to be experienced. That appears to be a severe requirement. If A-M assert that no *projected* values were used, they choose to ignore Krutilla's prepared testimony, with which as participants in the proceedings they were provided. In any event A-M provide their own projections (their Table 1), which incidentally are incorrect, but assuming that the degree of error is immaterial, their projected quantities underscore our position rather than undermine it. First of all the benefit per visitor day at $t=20$ while somewhat overstated, based on the assumptions they employ, is still a very modest figure. Estimates of the average benefit per fisherman day produced independently (Brown, Singh, and Edwards; and Mathews and Brown) for the mid-1960's exceed the value ($19.63) projected for the year 1996 by A-M. One would think that A-M would find comfort rather than fault with our results. Moreover, even the figure of $145.42 projected for the year 2026 compares rather modestly with present fees for prime Atlantic salmon fishing of $200–$500 per rod day in Norway, or even the $160 per rod day currently on prime beats for trout fishing on the River Test in Great Britain. What A-M appear to ignore is the rather substantial price inelasticity, and high income elasticity for prime outdoor recreation opportunities when the supply of such opportunities is rather scarce in relation to demand, as we are beginning to experience in the United States and which we can anticipate with increased severity by the year 2026. In this connection also, we would like to draw attention to the statement in our original paper (p. 616) that the projected implicit price or value of a recreation day in relation to projected per capita income in the terminal year of the evaluation horizon is roughly only *half* the ratio of the estimated current value to current per capita income.

In concluding, we must emphasize our view stated earlier that there is no basis for A-M's assertion of error in either our theoretical or empirical results.

REFERENCES

W. Brown, A. Singh, and E. Castle, "Net Economic Value of the Oregon Salmon-Steelhead Sport Fishery," *J. Wildlife Manage.*, Apr. 1965, *29*, 66–79.

———, ———, **and J. Edwards,** "Influence of Improved Estimating Techniques on Predicted Net Economic Values for Salmon and Steelhead," mimeo., Oregon State Univ., 1972.

M. Burns, "A Note on the Concept and Measure of Consumer's Surplus," *Amer. Econ. Rev.*, June 1973, *63*, 335–44.

O. Burt and D. Brewer, "Estimation of Net Social Benefits from Outdoor Recreation," *Econometrica*, Sept. 1971, *39*, 813–28.

C. Cicchetti, "A Multivariate Statistical Analysis of Wilderness Users in the United States," in J. Krutilla, ed., *Natural Environments: Studies in Theoretical and Applied Analysis*, Baltimore 1972.

———, A. Fisher, and V. K. Smith, "Evaluating the Opportunity Costs of Environmental Preservation: Mickey Mouse in the Mountains," mimeo., 1973.

A. Fisher, J. Krutilla, and C. Cicchetti, "The Economics of Environmental Preservation: A Theoretical and Empirical Analysis," *Amer. Econ. Rev.*, Sept. 1972, *62*, 605–19.

——— and ———, "Valuing Long-Run Ecological Consequences and Irreversibilities," paper presented at the Environmental Protection Agency Symposium on Cost-Benefit Analysis and Water Pollution Policy, Annapolis, Sept. 1973.

A. C. Harberger, "Three Basic Postulates for Applied Welfare Economics: An Interpretive Essay," *J. Econ. Lit.*, Sept. 1971, *9*, 785–97.

H. Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," *Econometrica*, July 1938, *6*, 242–69.

J. Krutilla, (1967a) "Conservation Reconsidered," *Amer. Econ. Rev.*, Sept. 1967, *57*, 777–86.

———, (1967b) *The Columbia River Treaty: The Economics of an International River Basin Development*, Baltimore 1967, 31–56.

S. Matthews and G. Brown, *Economic Evaluation of the 1967 Sport Salmon Fisheries of Washington*, Washington Dept. of Fisheries tech. rep., Olympia 1970.

E. Silberberg, "Duality and the Many Consumer's Surpluses," *Amer. Econ. Rev.*, Dec. 1972, *62*, 942–52.

G. Stankey, "A Strategy for the Definition and Management of Wilderness Quality," in J. Krutilla, ed., *Natural Environments: Studies in Theoretical and Applied Analysis*, Baltimore 1972.

Federal Power Commission (*FPC*), "In the Matter of: Pacific Northwest Power Company and Washington Public Power Supply System," Hearings, Washington 1970.

Gains from Trade Under Uncertainty

By Raveendra N. Batra and William R. Russell*

The theory of international trade has traditionally been built upon the assumption that the economic environment is characterized by the absence of uncertainty. But it is perhaps reasonable to expect that exposure of an economy to international trade contributes to uncertainty, whether or not the economic milieu is stochastic in the absence of trade. International trade introduces in the economy stochastic elements which are usually beyond the control of trading partners.

The purpose of this paper is to explore the implications of uncertainty for the welfare aspects of international trade theory. The first section of the paper is devoted to the development of the basic model and the derivation of the necessary conditions which ensure the maximization of expected utility for the society and the maximization of expected profits on the part of producers. In Sections II and III, we examine the implications of uncertainty about world prices for social welfare. Section IV is concerned with the question of optimal policy.

I. The Formal Model

Unless otherwise specified, the following assumptions will be maintained throughout the paper. Except for the modifications caused by the presence of uncertainty, our model is the usual two-country, two-good model of international trade.

1. There are two traded goods, X_1 and X_2, both of which are produced and consumed domestically. Furthermore, the first good is exported and the second good is imported.

2. Consumer preferences (actually the preferences of the representative consumer) are represented by a social utility function, $U=U(C_1, C_2)$, where C_i is the consumption of the ith good. There is no satiation in consumption, and the marginal utility of each good is decreasing so that $U_i>0$ and $U_{ii}<0$.

3. The transformation curve which is described by $X_2=X_2(X_1)$ is strictly concave to the origin. This implies that both X_2' and X_2'' are negative.

4. The country in question is small and is a price taker in the sense that it is unable to influence the probability distribution about the terms of trade p, where p is the world relative price of the first good in terms of the second good. Thus, the second good is taken to be the numeraire throughout; p is a random variable with a density function $f(p)$, expected value $E[p]=\mu$, and a probability distribution $F(p)$. Both producers and consumers make their economic decisions before p is known.

5. There is perfect competition in all markets, so that each producer seeking to maximize expected profits equates the marginal cost to the expected price.[1] Let $\overline{Y}$ stand for expected national income in terms of the second commodity. Then, with no impediments in trade,

$$\overline{Y} = E[p]X_1 + X_2 \tag{1}$$

Since producers are expected profit maximizers, they maximize the expected value of Y subject to the constraint provided by the transformation curve

$$X_2 = X_2(X_1) \tag{2}$$

* Professors of economics, Southern Methodist University. We are grateful to Josef Hadar and the managing editor for several useful suggestions.

[1] The assumption of expected profit maximization or risk neutrality on the part of producers has been made for two reasons. First, as it turns out, most of the results derived in this paper remain qualitatively unchanged if producers are risk averse. Second, in order to show the implications of risk aversion of producers for social welfare, we will have to construct a full-fledged production model incorporating factor markets as well. Space considerations do not allow for this exercise; furthermore, a pure two-sector production model has been developed under risk-averse behavior elsewhere by Batra (1974) and a similar presentation will be more or less a duplication.

The first-order condition for this constrained maximization is then given by

$$\frac{d\bar{Y}}{dX_1} = E[p + X_2'] = 0$$

whence

$$X_2' = -E[p] \tag{3}$$

Once the producers have made the optimal output decisions, the budget constraint for consumers given by

$$pC_1 + C_2 = pX_1 + X_2 \tag{4}$$

is determined. The social utility function is given by

$$U = U(C_1, C_2) \tag{5}$$

Since p is a random variable, utility is also random because randomness enters into the utility function when we substitute for C_2 from (4) to obtain

$$U = U(C_1, pX_1 + X_2 - pC_1)$$

Thus, the community maximizes

$$E[U] = E[U(C_1, pX_1 + X_2 - pC_1)] \tag{6}$$

by choosing C_1 before p is known, thereby treating C_2 as a random variable.[2] Differentiating (6) with respect to C_1, we obtain the first-order condition for a maximum:

$$\frac{dE[U]}{dC_1} = E[U_1 - pU_2] = 0 \tag{7}$$

In order to obtain further information from (7), it is necessary to specify how the expected values of U_2 and p are interrelated. If we assume that p and U_2 are independently distributed, a very strong assumption, $E[pU_2]$ equals $E[p]E[U_2]$, so that from (7) $E[U_1]/E[U_2]=E[p]$. In this case, the consumers maximize expected utility by equating the expected value of p to the ratio of expected marginal utilities or the marginal rate of substitution. However, the independence assumption is overly strong, because in general, U_2 and p are correlated. Furthermore, this correlation is negative, because a lower value of p in the absence of inferiority in social consumption will be associated with a lower consumption of the second good and thus a higher level of U_2, thanks to diminishing marginal utilities. Let $\sigma=E[pU_2]-E[p]E[U_2]$ be the covariance of p and U_2. In view of the negative relationship between p and U_2, $\sigma<0$. Substituting this in (7), we obtain

$$\frac{E[U_1]}{E[U_2]} = E[p] + \frac{\sigma}{E[U_2]} \tag{8}$$

So long as σ is nonzero, $E[p]$ differs from the ratio of expected marginal utilities and hence the marginal rate of expected substitution, as the community maximizes its expected utility.

With equation (8), the development of our model is complete. The main difference between our model and the traditional certainty models lies in the fact that under certainty, profit maximizing producers equate X_2' to $-p$ and the utility maximizing society equates U_1/U_2 to p, whereas under conditions of uncertainty about the terms of trade, the necessary maximizing conditions are given by (3) and (8).

II. Welfare Implications of Increased Uncertainty

This section is concerned with the implications of a change in the probability distribution of p for expected social welfare.

Let p have the density function $f(p)$. From equation (4) we obtain the density $g(C_2)$ and likewise the distribution function $G(C_2)$ for the random variable C_2 as

[2] It is interesting to note that our formulation of the problem is relevant to several developing countries. Many of these countries face uncertainty about their terms of trade either because of the past fluctuation in the world prices of their exportables or because of the nature of the goods they export. Quite often, the central planners or the architects of commercial policy seek to fulfill certain export targets before the knowledge of actual world prices either in order to alleviate the critical foreign exchange shortage or in order to obtain capital goods from abroad for domestic investment. However, setting the export target in advance of the resolution of the terms of trade is equivalent to choosing the level of C_1 for the community, because the optimal output of X_1 is determined by the producers.

$$G(C_2) = F\left(\frac{C_2 - X_2}{X_1 - C_1}\right)$$

The range of C_2 is bounded by:

$$\text{lower bound for } C_2 = \lim_{p\to 0}\,[X_2 + p(X_1 - C_1)] = X_2$$

$$\text{upper bound for } C_2 = \lim_{p\to p_0}\,[X_2 + p(X_1 - C_1)] = X_2 + p_0(X_1 - C_1) \equiv b$$

where p_0 is the upper bound for p. Expected utility is then written as

$$E[U] = \int_{X_2}^{b} U(C_1, C_2)g(C_2)dC_2$$

We are now in a position to evaluate the change in expected utility due to a change in the density function for the terms of trade. In particular, we are interested in the sign of the change in the expected utility as the density function of C_2 induced by a change in the density function of p changes from $g_1(C_2)$ to $g_2(C_2)$, or in the sign of

$$(9) \qquad \Delta E[U] = \int_{X_2}^{b} U(C_1, C_2) \cdot [g_2(C_2) - g_1(C_2)]dC_2$$

where $f_1(p)$, $g_1(C_2)$ are the density functions depicting the original expectations and $f_2(p)$, $g_2(C_2)$ are the new ones. It may be noted here that since C_2 from (4) is a linear transformation of p, a change in the probability distribution of p from F_1 to F_2 induces a change in the C_2 distribution from G_1 to G_2.

Integrating (9) by parts, we obtain

$$(10) \qquad \Delta E[U] = [U(C_1, C_2)\{(G_2(C_2) - G_1(C_2)\}]_{X_2}^{b} - \int_{X_2}^{b} U_2(C_1, C_2)[G_2(C_2) - G_1(C_2)]dC_2$$

But $G_2(X_2) - G_1(X_2) = 0 = G_2(b) - G_1(b)$ since X_2 and b are the lower and upper bounds, respectively, of the ranges of both distributions. Hence, the first term in equation (10) is zero. Since $U_2(C_1, C_2) > 0$, expected utility decreases if

$$(11) \qquad G_2(C_2) \geq G_1(C_2)$$

for all C_2, $X_2 \leq C_2 \leq b$, and the strict inequality of the conditions holding at least once. In the literature this condition is called first-degree stochastic dominance. As is well known, the condition simply amounts to asserting that when the two cumulative distributions for C_2, $G_2(C_2)$, and $G_1(C_1)$ are plotted, then if G_1 lies on or below the G_2 distribution and in at least some range lies strictly below, then we say that G_1 stochastically dominates G_2 in the first degree,[3] or symbolically, G_1FSDG_2.

In the context of our problem, we can say unambiguously that if there is a change in the distribution of C_2 from G_1 to G_2 and G_1FSDG_2, then expected utility declines as a result of this change in distribution. It may be noted that the only condition that is needed to determine the direction of the change in expected utility is that the marginal utility of consumption of the second good is positive. Therefore, with the *FSD* type of change in the probability distribution, the expected utility declines as a result of the shift in the distribution from G_2 to G_1 even if the utility function is not risk averse.

Other more interesting changes in distributions will also yield determinate results. To show this, we first require an alternative expression for the change in expected utility. Integrating (10) by parts we have

$$(12) \qquad \Delta E[U] = -\left[U_2(C_1, C_2)\left(\int_{X_2}^{C_2} G_2(t)dt - \int_{X_2}^{C_2} G_1(t)dt\right)\right]_{X_2}^{b} + \int_{X_2}^{b} U_{22}(C_1, C_2) \cdot \left[\int_{X_2}^{C_2} G_2(t)dt - \int_{X_2}^{C_2} G_1(t)dt\right] dC_2$$

With $U_2 > 0$ and $U_{22} \leq 0$, it is clear that

[3] See, for example, James Quirk and Rubin Saposnik, Haim Levy and M. Sarnat, and Josef Hadar and Russell (1969). The concepts of stochastic dominance have been further generalized in Hadar and Russell (1971; 1974).

$\Delta E[U]<0$ if

$$(13) \qquad \int_{X_2}^{C_2} G_2(t)dt \geq \int_{X_2}^{C_2} G_1(t)dt$$

for all C_2 with the strict inequality holding at least at some point. This condition is called second-degree stochastic dominance. We may alternatively write this condition as G_1SSDG_2 and read G_1 stochastically dominates G_2 in the second degree.

As with the *FSD* case, we can geometrically describe the *SSD* condition between the two distributions; G_1 stochastically dominates G_2 in the second degree if, and only if, for every point in the range of the random variable, the area under the distribution G_1 is equal to or less than that of G_2, with G_1 area being strictly less at least in some interval.

Unlike the earlier *FSD* case, the direction of the change in expected utility as a result of the change in the distribution from G_1 to G_2, with G_1SSDG_2 cannot be determined just from the strict monotonicity of the utility function. This is because the sign of $\Delta E[U]$ from (12) is determined not only by the sign of U_2 but also by the sign of U_{22}. Risk aversion of the community is sufficient to ensure the ranking of the probability distributions by expected utility, given that they can be ordered by the *SSD* relation. In other words, given our assumption that $U_{22}<0$, if the original distribution of C_2 stochastically dominates the new distribution, then there will be a decline in expected social utility. It should be noted again that either *FSD* or *SSD* condition on F_1 and F_2 is preserved in G_1 and G_2.

Up to now we have examined the implications of a change in the probability distribution of p for expected social welfare. We will now show that the expected utility effects of an increase in uncertainty about the terms of trade can be obtained as a special case of the results derived above.

An increase in risk or uncertainty is sometimes defined as a "stretching" of the density function around a constant mean. That is, a new riskier distribution can be obtained by a procedure which spreads the original density function to induce greater area under the function for regions farther away from the mean while keeping the mean invariant. This procedure of spreading the distribution while preserving the mean, however, yields the condition F_1SSDF_2 where F_2 is the distribution obtained by spreading F_1.

If we restrict our discussion to those distributions which have equal means, equation (12) reduces to a simpler form. It can be shown that

$$\int_{X_2}^{b} G_2(C_2)dC_2 - \int_{X_2}^{b} G_1(C_2)dC_2 = \mu_1 - \mu_2$$

Under the mean-preserving spread, the first term in (12) yields[4]

$$-\left[U_2(C_1, C_2)\left(\int_{X_2}^{C_2} G_2(t)dt - \int_{X_2}^{C_2} G_1(t)dt\right)\right]_{X_2}^{b} = U_2(C_1, b)(\mu_1-\mu_2)=0$$

As stated above, it is popular to interpret the distribution with the larger spread but constant mean as being "more risky" than the original distribution. But since we know that if these distributions are G_2 and G_1, respectively, then G_1SSDG_2. From (12) then, we see immediately that $\Delta E[U]<0$ provided, of course, that $U_{22}<0$ and the society is risk averse. Hence, an increase in risk or uncertainty leads to a decline in expected social utility.

As another special case, we may consider

[4] In general, for nonnegative random variables,

$$\mu = \int_0^\infty xf(x)dx = \int_0^\infty f(x)\int_0^x dtdx = \int_0^\infty \left[\int_t^\infty f(x)dx\right]dt$$

$$\int_0^\infty [F(\infty) - F(t)]dt = \int_0^\infty [1 - F(t)]dt$$

Therefore,

$$\mu_1 - \mu_2 = \int_0^\infty [1 - F_1(t)]dt - \int_0^\infty [1 - F_2(t)]dt = \int_0^\infty [F_2(t) - F_1(t)]dt$$

the comparison between the level of social utility attainable under certainty with that attainable under uncertainty. There is no obvious way of making this comparison because under certainty, $-X_2'=p=U_1/U_2$ whereas under uncertainty, the equivalent condition is given by

$$-X_2' = E[p] = \frac{E[U_1]}{E[U_2]} - \frac{\sigma}{E[U_2]}$$

But there is one reasonable way of formulating the problem: What is the level of social welfare under uncertainty as compared to the case where the terms of trade are known with certainty to be equal to the mean of the original distribution? In other words, we restrict $E[p]$ to equal μ, the certainty terms of trade. Notice that whatever the distribution of p in the uncertain case $(F_2(p))$, it entails

$$F_1(p) - F_2(p) \leq 0 \qquad \text{for } 0 \leq p \leq \mu$$

and

$$F_1(p) - F_2(p) \geq 0 \qquad \text{for } \mu \leq p \leq p_0$$

where for the certainty distribution $F_1(p)=0$ for $p<\mu$ and $F_1(p)=1$ for $p\geq\mu$. With equal means

$$\int_0^{p_0} [F_1(\mathrm{t}) - F_2(\mathrm{t})]d\mathrm{t} = 0$$

From this, it follows that $\int_0^p[F_1(\mathrm{t})-F_2(\mathrm{t})]d\mathrm{t} \leq 0$ for all p, $0\leq p\leq p_0$. Thus, the stochastic dominance condition (13) holds. Hence, from equation (12) we conclude that the expected utility declines as the probability distribution of p changes from that of the certain case to that of uncertainty with the same mean. We can also see from equation (8) that the choice of C_1 would be such that $E[p]=E[U_1]/E[U_2]-\sigma E[U_2]$ and hence the utility of C_2 exceeds the expected utility of C_2. Thus, even if the actual p turns out to be the p in the certain case, the choice will not be a point of maximum utility. The loss will be the cost of hedging against possible alternative outcomes by weighting losses more heavily than equivalent gains in the risk-averse welfare function.

III. Geometrical Exposition

The geometrical exposition of the results derived in the previous sections turns out to be very rewarding in terms of clarity and comprehension. As a prelude to later discussion in this section, we first present the model where the world terms of trade are unknown to producers as well as consumers but all decision makers are absolutely certain about what prices are going to be and act with full confidence on this expectation.

We assume here for simplicity that producers and consumers (actually the representative consumer) have the same expectations about world prices. If economic agents act with certainty about future prices, then profit-maximizing producers will equate their marginal rate of transformation to the negative of the expected world price ratio and produce at a point such as P in Figure 1, where $\overline{FP}$, whose slope indicates the expected international relative price of the first good in terms of the second, touches the transformation curve TT' at P. Similarly, consumers acting with certainty about the expected price ratio will maximize their expected utility by consuming at a point such

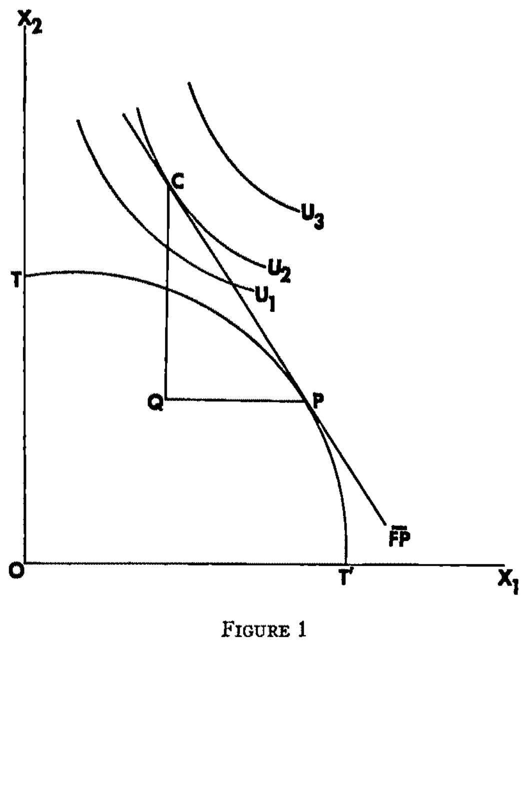


FIGURE 1

as C where the social indifference curve $\overline{U}$ touches the budget line $\overline{FP}$. Given a set of social indifference curves such as $\overline{U}_1$, $\overline{U}_2$, $\overline{U}_3$, etc., the community chooses that indifference curve which is tangential to the expected price line, so that the marginal rate of substitution is equated to the expected price ratio. This selection of consumption and production points also means that the society expects to export PQ of X_1 in exchange for CQ of X_2 from abroad.

The general case is one where the decision makers take into account the fact that they could be "wrong" about the future prices. They recognize that the terms of trade could be higher or lower than those expected and hence they hedge against the possibility of being wrong. The assumption that producers are expected profit maximizers implies risk neutrality on their part. Therefore, they continue to equate the marginal rate of transformation to $-E[P]$. However, since the consumers are risk averse, from (8) they equate the marginal rate of substitution not to $E[p]$ but to $E[p]+\sigma/E[U_2]$. In other words, they attach greater subjective weight to the possibility of losses which could

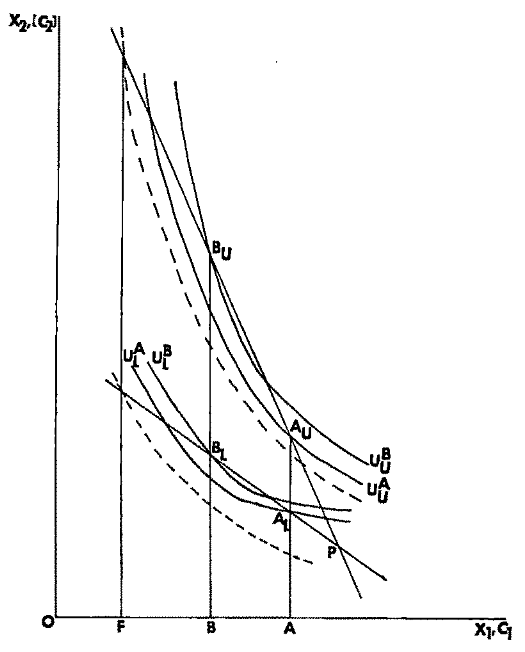


FIGURE 2

occur if p turned out to be lower than $E[p]$, than to the windfall gains which could occur if p turned out to be higher than $E[p]$.

Let us now consider Figure 2, which is constructed in the light of this discussion. As in Figure 1, P is again the production point, which is obtained from the tangency of the $\overline{FP}$ line to the transformation curve (neither of which is shown in Figure 2 in order to avoid cluttering the diagram). Let p_L be the lower bound for p and p_U be the upper bound. The consumers facing a given probability distribution of p realize that p can take any value from p_L to p_U with the mean of the distribution being $E[p]$. Since p can take any value from p_L to p_U, this means that the budget constraint facing the consumers can take any value from $p_LX_1+X_2$ to $p_UX_1+X_2$. Faced with this randomness of income, consumers make that choice of C_1 which maximizes their expected utility. Suppose C_1 is chosen to be equal to OA; then corresponding to this choice of C_1, the lower bound for C_2 is given by A_L and its upper bound by A_U. In other words, the choice of C_1 generates a certain probability distribution for C_2, which in turn depends on the probability distribution of p. In general, with a continuous density function for p, all values of C_2 between A_L and A_U are available when C_1 is chosen to be OA. If C_1 were chosen at point B, the lower and upper bounds for C_2 would be given by B_L and B_U, respectively. Therefore, each choice of C_1 generates a certain distribution of C_2, which in turn generates a certain distribution of utility. For example, the choice of C_1 at A yields a lower bound of utility of U_L^A and an upper bound of U_U^A and similarly with U_L^B and U_U^B. The consumer makes that choice of C_1 which generates a utility distribution with the highest mean or $E[U]$. For example, comparison of the utility distributions associated with the two points A and B shows that utility attainable at both the upper and lower bounds is higher with B than with A. Clearly, when B is chosen $E[U]$ is higher than the $E[U]$ when A is chosen. Similarly, we could perhaps find another choice of C_1 which will be even superior to point B. In general, the optimal choice of C_1 is deter-

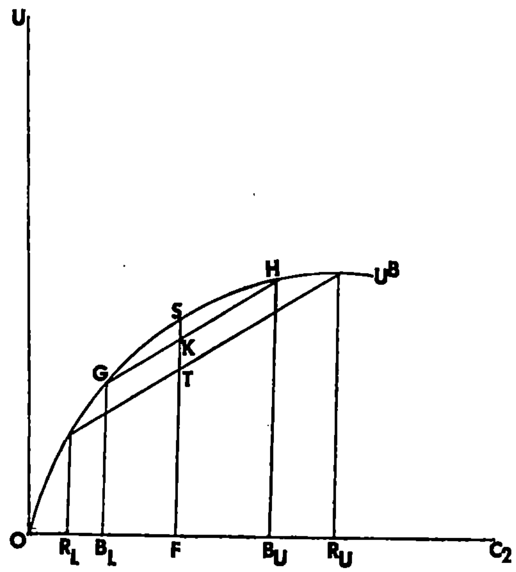


FIGURE 3

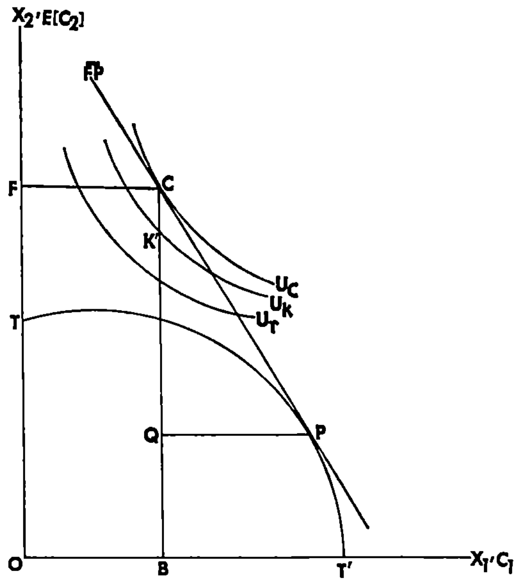


FIGURE 4

mined by the shape of the utility map and the probability distribution of p. In Figure 2, as less of C_1 is consumed, $E[U]$ rises, but this need not be true of all points to the left of A. Point F, for example, generates upper and lower bounds of utility levels that are given by the dashed social indifference curves which clearly lie below the respective upper and lower bounds of utility associated with point A.

Where then is the equilibrium consumption point, and what is the expected utility level associated with it? The answer, using the discrete two-state distribution case, is illustrated by Figures 3 and 4. Figure 3 is the classic Friedman-Savage diagram where utility is plotted against the random variable C_2, given that the choice of C_1 has already been made at, say, point B. The shape of the curve OU^B shows that the marginal utility of the second good declines as C_2 increases; that is, the consumers are risk averters. The lower and upper bounds for C_2 are given by B_L and B_U. Since expected utility is a linear combination of the utility levels attainable at B_L and B_U, the level of utility corresponding to any point between B_L and B_U is given by a point on the straight line GH. Figure 4, on the other hand, is similar to the trade diagram given by Figure 1; P is the production point as before, but lines p_L and p_U are not drawn. The choice of C_1 at B means that the expected consumption point will lie at C so that the expected budget constraint, which in turn is determined by the actions of the producers, is satisfied. At C, the level of X_2 consumed is given by OF, and corresponding to this level of C_2, the expected utility in Figure 3 is given by FK. If there were no uncertainty in the economy, then for the same level of $E[C_2]$ and hence $E[p]$, the utility level would be given by FS. This shows that social welfare under certainty is higher than that available with uncertainty. In terms of Figure 4, the level of utility under certainty is given by U_c but the expected utility level is given by U_k, and points K' and C correspond respectively to points K and S in Figure 3.

Several points deserve further consideration. First, by consuming at C, the country chooses to export PQ of X_1 and import CQ of X_2. Second, point K' is such that the necessary condition for expected utility maximization, condition (8), is satisfied. At C, the marginal rate of substitution equals $-E[p]$, whereas at K', it equals $-E[U_1]/$

$E[U_2]$ and the difference between the slopes of the social indifference curves at points K' and C is given by $-\sigma/E[U_2]$.

Suppose there is an increase in uncertainty of the mean-preserving type. This means that for the same choice of C_1 and hence $E[C_2]$, the lower and upper bounds for C_2 move further apart, say to R_L and R_U in Figure 3. It is evident then that for the same mean of C_2 at F, expected utility declines from FK to FT. In terms of Figure 4, this decline is portrayed by the shift from U_k to U_r.

IV. Optimal Policy Under Uncertainty

Until now, our analysis has been based on the fact that there are no impediments in the free flow of goods across the nations. In this section, we wish to examine whether free trade, in fact, is the optimal policy for a small country. Our method of analysis here corresponds to that typically followed by trade theorists such as Murray Kemp, Ronald Jones, Roy Ruffin, and Batra (1973) in analyzing the questions of optimal policy under certainty. We define optimal policy to be one which maximizes the expected value of social utility subject to the constraints provided by the transformation curve and the budget. Assuming an initial situation of free trade, the latter will be the optimal policy only if $dE[U]=0$, because no other policy will then be able to effect an increase in expected utility. However, if $dE[U]\neq 0$, then that alternative policy which fulfills the necessary condition for maximization will be the optimal policy.

When world prices are known with certainty, the profit-maximizing behavior of the producers and the utility-maximizing behavior of the consumers along with our other assumptions specified at the beginning of Section I are sufficient to ensure the optimality of free trade. This is because, since p is nonrandom, $E[p]$ is simply p and $\sigma=0$. The question now arises whether the expected profit maximization and expected utility maximization constrained by the actions of producers will yield the highest level of expected utility.

Differentiating the expected utility function given by (6) totally and using (7), we obtain

$$dE[U] = dX_1(E[U_1] + X_2' E[U_2])$$

Now if p and U_2 are independent, then $dE[U]=0$, in which case the initial situation of free trade is optimal. This is because

$$\frac{E[U_1]}{E[U_2]} = E[p] = -X_2'$$

However, as argued before, this is a very strong assumption which is not generally satisfied. In the normal case, p and U_2 are negatively correlated and, therefore, $dE[U]\neq 0$, so that free trade is not the optimal policy. To discover the optimal policy, let us equate $dE[U]$ to zero to obtain

$$\frac{E[U_1]}{E[U_2]} = -X_2' \tag{14}$$

Since $X_2'=-E[p]$, optimality calls for equating the marginal rate of substitution to the expected price ratio. The logic behind this policy should be apparent in view of our analysis in the previous section. We have already established that the presence of uncertainty causes a decline in expected social utility. The maximum level of expected utility attainable in Figure 4 is given by U_c, which is obtained when the marginal rate of substitution equals $-E[p]$. However, U_c cannot be attained so long as world prices are uncertain and consumers are risk averse. Hence, optimality calls for the elimination of uncertainty to consumers. This can be accomplished if the government assures the consumers of the stability of future prices at the level furnished by the mean of the probability distribution of p. In other words, the government undertakes to subsidize (or tax) the consumption of the traded goods in case the actual terms of trade turn out to be different from the expected ones.

There is yet another way of achieving the optimum. Since $X_2'=-E[p]$, equation (14) will be satisfied if $E[U_1]/E[U_2]$ is equated to $E[p]$. From equation (8), with $\sigma<0$, $E[U_1]/E[U_2]$ is less than $E[p]$. Therefore, optimal policy calls for lowering the consumption of C_1 and raising that of C_2 until

$E[U_1]/E[U_2]$ is raised to the level of $E[p]$. A consumption subsidy to the second good and/or a consumption tax on the first good will achieve the desired result.

The economic logic behind this policy is this: Consumers choosing C_1 before p is known tend to overconsume the first good and thus attach a lower value of $E[U_1]$ to it than the producers, who are not risk averters. This causes a difference between the marginal rate of transformation and the marginal rate of expected substitution. The consumption of the first good should then be lowered to eliminate this difference, thereby bringing the producers' relative valuation of goods in line with the consumers' relative valuation.

REFERENCES

R. N. Batra, "Resource Allocation in a General Equilibrium Model of Production Under Uncertainty," *J. Econ. Theory*, May 1974, *8*, 50–63.

———, *Studies in the Pure Theory of International Trade*, London 1973.

J. Hadar and W. R. Russell, "Rules for Ordering Uncertain Prospects," *Amer. Econ. Rev.*, Mar. 1969, *59*, 25–34.

——— and ———, "Stochastic Dominance and Diversification," *J. Econ. Theory*, Sept. 1971, *3*, 288–305.

——— and ———, "Diversification of Interdependent Prospects," *J. Econ. Theory*, Mar. 1974, *7*, 231–40.

R. W. Jones, "International Capital Movements and the Theory of Tariffs and Trade," *Quart. J. Econ.*, Feb. 1967, *81*, 1–38.

M. C. Kemp, "The Gain from International Trade and Investment: A Neo-Heckscher-Ohlin Approach," *Amer. Econ. Rev.*, Sept. 1966, *56*, 788–809.

H. Levy and M. Sarnat, "A Note on Portfolio Selection and Investors' Wealth," *J. Finance Quant. Anal.*, Jan. 1971, *6*, 639–42.

J. P. Quirk and R. Saposnik, "Admissibility and Measurability Utility Functions," *Rev. Econ. Stud.*, Feb. 1962, *29*, 140–46.

R. J. Ruffin, "Tariffs, Intermediate Goods, and Domestic Protection," *Amer. Econ. Rev.*, June 1969, *59*, 261–69.

Technological Change, Production, and Investment in Natural Resource Industries

By GORDON C. RAUSSER*

In recent years, the possibility of alleviating the apparent inadequacy of nature's endowments by rapid technical change has become increasingly obvious. Some have argued that technical progress has sufficiently expanded the resource base so that the scarcity forecasted by Malthus, for example, has not arisen nor is it likely that it will arise. To support this view, most authors refer to available statistical evidence which suggests that the (relative) direct costs of production from natural resources have fallen over time (see John Krutilla). A similar position is assumed by a number of public agencies and congressional committees.[1]

Theoretical contributions have been advanced in a natural resource context which are motivated, in part, by the recognition that technical progress has altered the scarcity view of Malthus. However, these contributions have treated technical knowledge as an exogenous rather than an endogenous component. The most general of these treatments are the works of Oscar Burt and Ronald Cummings, and Vernon Smith.[2] Although these authors explicitly introduce capital investment into a production model for natural resources, they unfortunately treat technical progress as manna from heaven. The most comprehensive of these models, developed by Burt and Cummings, takes technology into account by specifying the criterion function (social benefit function) of natural development as an explicit function of time. In the examination of convergence and existence conditions for an infinite intertemporal optimization problem, however, they assume that the social benefit function is not explicitly a function of time.[3] This apparent inconsistency can be easily removed by treating technical knowledge augmentations as an endogenous component of a general natural resource model. Such a model would not be subject to the major limitation noted by the above authors in an earlier paper, viz., "any industry model would be very deficient without cognizance of the major influence of technical change" (Cummings and Burt, p. 988).

It is desirable to distinguish between two types of technical progress. The first results from research and innovation activity while the second results from experience in production, investment, etc., i.e., "learning by doing."[4] In this paper I shall be concerned only with the learning by doing type of technical progress. Given that many investments in natural resource industries are often (subjectively) justified on the basis of learning by doing, a comprehensive model

* Associate professor of economics and associate agricultural economist in the Experiment Station and on the Giannini Foundation, University of California, Davis.
Research support by the University of California Water Resource Center under project W-333 is gratefully acknowledged. Giannini Foundation research paper 374.

[1] For example, legislation will be introduced shortly in the U.S. Senate calling for a $20 billion, ten-year effort in the hope of achieving a self-sufficiency state for *U.S.* energy needs. This legislation, based on a detailed study by the Interior Committee, refers specifically to the development of "energy technologies" which will allow the utilization of *U.S.* natural resources to meet demand by 1983.

[2] A number of specific resource models are surveyed in Burt and Cummings.

[3] As these authors lucidly argue, this assumption is not particularly restrictive once we recognize that projections of the future often "... become so nebulous after about 50 years that the best guess is to assume ..." (p. 589) the social benefit function is constant at some point (finite) in the future after which potential changes are, to say the least, vague.

[4] For both types of technical progress there is a relatively large and growing literature. Much of this literature is summarized in my 1973 paper. For an indication of the empirical importance of the learning by doing type of technical progress in the context of a particular natural resource, see the author, C. Willis, and P. Frick.

which explicitly incorporates this dimension, simultaneously optimizing the rate of resource use and investment, would appear to be of some value. Such a model would indicate, for example, whether or not atomic power plants (whose present costs exceed those of alternative energy sources) ought to be adopted. This decision would presumably be determined on the basis of the knowledge gained in their construction and operation relative to the additional cost per kilowatt hour of this energy source. A host of other possible examples are not difficult to isolate.

To maintain some continuity with previous theoretical contributions, I shall proceed by modifying and extending the treatment contained in Burt and Cummings for natural resource industries. The intertemporal optimization problem underlying the model will be viewed from the standpoint of society. Technical knowledge for various processes will be introduced as an endogenous component. More specifically, the model development will evolve by specifying in Section I a framework for social internal learning, production, and investment in natural resource industries. Economic interpretations of the resulting optimum time path are treated in Section II along with equilibrium conditions under an infinite planning horizon. These interpretations are compared to the Burt and Cummings result based on exogenous technical progress, and the advantages of the present formulation are indicated.

I. Model

A societal planning horizon beginning with period $t=0$ and ending at period $t=T$ will be assumed. The amount of resource stock available at t will be denoted by x_t while the level of capital stock is designated by y_t. Associated with the resource stock is a rate of resource utilization, production, or extraction at time t, u_t, and associated with capital stock is gross investment v_t at time t. In addition to the two stock variables x_t and y_t, and the two flow variables u_t and v_t, we require variables representing accumulated knowledge at various stages of the planning horizon. These knowledge stocks will be specified for the resource utilization process, resource stock developments, and the capital investment process. That is, define W_t^u as the stock of knowledge pertaining to resource production at the beginning of period t, W_t^x as the accumulated knowledge associated with resource stocks at the beginning of period t, and W_t^y as the accumulated knowledge related to capital investment and thus stocks at the beginning of period t.

The rate of resource use is assumed to satisfy the constraint

$$(1) \quad u_t \leq h(x_t, y_t, W_t^u), \quad t = 0, \ldots, T-1$$

where $h(x_t, y_t, W_t^u)$ is a quasi-concave function. The relationship $h(\ )$ is such that larger resource, capital, and production knowledge stocks imply larger rates of production are allowable, i.e., $\partial h/\partial y_t$, $\partial h/\partial x_t$, $\partial h/\partial W_t^u \geq 0$. For some applications u_t will be more appropriately viewed as a rate of production, for others as a rate of extraction, and for still others simply as a rate of utilization. As the stock of knowledge associated with this rate (W_t^u) is increased, the efficiency with which say minerals are extracted, gas is produced, or groundwater is utilized, is improved.

At the beginning of any period $t+1$, resource, capital, and knowledge stocks are presumed to be obtained from the difference equations:

$$(2) \quad x_{t+1} = x_t + g(u_t, x_t, W_t^x)$$

$$(3) \quad y_{t+1} = y_t - D(u_t, v_t, y_t, W_t^y)$$

$$(4) \quad W_{t+1}^x = W_t^x + \omega^x(u_t, W_t^x)$$

$$(5) \quad W_{t+1}^y = W_t^y + \omega^y(v_t, W_t^y)$$

$$(6) \quad W_{t+1}^u = W_t^u + \omega^u(u_t, W_t^u)$$

$$t = 0, 1, \ldots, T-1$$

The functions $g(\ )$, $\omega^k(\ )$, $(k=x, y, u)$ are each assumed quasi concave and the function $D(\ )$ is quasi convex. The function $g(\ )$ obviously represents the net additions to resource stocks during period t; $D(\ )$ represents a depreciation function which measures the net changes in capital stocks during period t; and the functions $\omega^k(\ )$ represent

the net additions to knowledge associated with each process k; i.e., they represent the net learning functions related to the x, y, and u processes.

The assumed properties of the relationship $g(\ )$ are: $\partial g/\partial x_t \geq 0$, $\partial g/\partial u_t \leq 0$, and $\partial g/\partial W_t^x \geq 0$; i.e., (relatively) larger resource stocks, smaller rates of utilization, and larger knowledge stocks associated with resource availability result in greater additions to resource stocks. In the case of groundwater, the inclusion of W_t^x in $g(\ )$ implies that recharges of aquifers less amounts pumped are more efficiently accomplished the greater the learning experience. A similar interpretation holds for the fossil fuel industry where $g(\ )$ might be specified as the difference between discoveries and the quantity produced or extracted during period t. For relationship $D(\ )$, depreciation is specified to be a nondecreasing function of the rate of resource utilization as well as the level of capital stocks, and a nonincreasing function of the rate of gross investment along with the level of knowledge stocks related to capital accumulation; i.e., $\partial D/\partial u_t \geq 0$, $\partial D/\partial y_t \geq 0$, $\partial D/\partial v_t \leq 0$, and $\partial D/\partial W_t^y \leq 0$. The presence of W_t^y in $D(\ )$ implies that as learning experience associated with the capital investment process increases, the rate of depreciation declines while the presence of u_t in the same function ". . . measures the consumption of capital associated with current production" (Burt and Cummings, p. 580).

The difference equations (4), (5), and (6) which specify the evolvement of knowledge for various processes over time are a generalization of the usual progress functions found in the learning by doing literature. More specifically, the learning function $\omega^k(\ )$ includes as its arguments not only the rate of production or investment but also the stock of knowledge in the current period. It is assumed that $\partial\omega^k/\partial W_t^k < 0$, and the remaining argument of each learning function is presumed to have a nonnegative effect, i.e., $\partial\omega^k/\partial u_t \geq 0$, for $k=x$, u and $\partial\omega^k/\partial v_t \geq 0$, for $k=y$, and each of the $\omega^k(\ )$ is presumed to be quasi concave. A simple representation of W^k results when knowledge depreciates at a constant rate, say α, $0<\alpha<1$. Such a specification in contrast to the usual progress function formulation would not lead to W^k being generated by a simple sum of past utilizations or investments, rather W^k would be represented by a weighted sum of these past realizations.[5]

Note that the general specifications for the stocks of knowledge (4), (5), and (6) explicitly allow only for internal learning. As these stocks are scalar values and refer specifically to a particular natural resource industry, knowledge may be treated as completely vested in the decision maker(s) or resource industry in question and hence nontransferable. Moreover, it is obvious that these learning function specifications treat technology advancements of the learning by doing type as endogenous. In contrast to the Burt and Cummings model, the functions $h(\ )$, $g(\ )$, and $D(\ )$ or the benefit function (reflecting fewer inputs to achieve the same rate of utilization, resource, and/or capital levels over time) are not regarded as exogenously influenced by technical progress, the latter being somehow determined by outside forces. Rather the functions $h(\ )$, $g(\ )$, and $D(\ )$ as indicated in (1), (2), and (3) are determined in part by the levels of technical knowledge which in turn are endogenously generated by (4), (5), and (6).

Any optimal selection of the rate of utilization (u_t) and the rate of investment (v_t) must, of course, satisfy the difference equations (1) through (6). These equations represent constraints to a societal optimization problem for a particular natural resource industry where the stock values x_{t+1}, y_{t+1}, and W_{t+1}^k may be viewed as state variables, and the flow values u_t and v_t may be viewed as control or decision variables. In addition to these constraints, all control and state variables are restricted to nonnegative values. That is,

$$(7) \qquad u_t,\ v_t,\ x_t,\ y_t,\ W_t^x,\ W_t^y,\ W_t^u \geq 0$$

[5] To illustrate, consider W_t^u, i.e., (6). For the constant rate of depreciation α and proportional learning γ specification, $\omega^u(\ )=\gamma u_t - \alpha W_t$ and

$$W_t^u = (1-\alpha)^t W_0^u + \gamma \sum_{j=1}^{t} (1-\alpha)^{j-1} u_{t-j}$$

for all t. Furthermore, the selection of the optimal path for rates of resource recovery and addition to capital stock and knowledge stock augmentations is constrained by initial conditions on each of the state variables. In other words, at $t=0$, the set of state variables are known constants, i.e.,

$$(8) \qquad x_0 = \hat{x}, \qquad y_0 = \hat{y}, \qquad W_0^x = \hat{W}^x, \qquad W_0^y = \hat{W}^y, \qquad W_0^u = \hat{W}^u$$

To complete the specification of the societal optimization problem for the use of a particular natural resource, we require a net (social) benefit function by which alternative time paths of the state and control variables are evaluated. Such a function is presumed to exist and to be quasi concave. In general form, this social benefit function will be designated as $B_t(u_t, v_t, x_t, y_t, W_t^x, W_t^y, W_t^u)$. Note that the subscript t on the benefit function is introduced to reflect in part the influence of advancing technology which is exogenous in nature. This form of technical change along with the learning by doing endogenous type of technical progress depicted in (4) through (6) is assumed to represent total technological advancement. It is further assumed that society desires to maximize the present value of these benefits over the time interval [0, T]. For the case in which the planning horizon is finite, i.e., $T<\infty$, a quasi-concave terminal value function, $\psi(x_T, y_T, W_T^x, W_T^y, W_T^u)$, will be specified so as to establish some continuity with future periods beyond T.

The rationale for including the arguments u_t, v_t, x_t, and y_t in the benefit function $B(\ )$ has been discussed by Burt and Cummings. Briefly, u_t is included since resource use presumably has some value in economic production or as a consumption (direct) good; v_t is included to reflect its opportunity costs in the remainder of the economy; x_t is included to reflect accessibility of the resource (for example, large stocks of coal are associated with relatively accessible and rich deposits); and y_t is included to reflect the possibility of lower costs of production per period associated with higher stocks of capital. The basis for incorporating the knowledge stocks W^k, $k=x, y, u$, might emanate from their impact on production functions implicit to $B(\ )$. Their inclusion could result also from the psychic (education or consumption) effects of acquiring additional knowledge. More concretely, they might appear as arguments of $B(\ )$ and $\psi(\ )$ to reflect the possibility of renting the services of generated knowledge elsewhere. Another possible rationale could result from decision makers' placing some value on transferring knowledge to public decision makers concerned with other natural resource industries. If this latter basis is relevant for a particular natural resource industry, then learning is not entirely specific or internal.

Summarizing the above discussion more formally, the postulated societal optimization problem for a particular natural resource industry may be stated as

$$(9) \quad \text{Max}\left[\sum_{t=0}^{T-1} \beta_t(u_t, v_t, x_t, y_t, W_t^w, W_t^y, W_t^u)\beta^t + \psi(x_T, y_T, W_T^x, W_T^y, W_T^u)\beta^T\right]$$

subject to (1) through (8) where the variables of optimization are u_t, v_t, $t=0, 1, \ldots, T-1$, and x_t, y_t, W_t^x, W_t^y, W_t^u, $t=1, \ldots, T$; T is assumed finite; Max denotes the maximum operator; and $\beta=1/(1+r)$, r being the appropriate positive discount rate.[6]

Assuming the usual limiting and continuity properties for the functions (1) through (9), the above optimization problem can be treated within a Kuhn-Tucker analytical framework. Assuming each state variable is positive over all periods of the planning horizon,[7] the necessary conditions for the decision variables u_t and v_t are dis-

[6] Since a single natural resource is assumed, the analysis is clearly a partial one and thus the discount rate should be based on an opportunity cost of capital measure to the remainder of the economy.

[7] A mathematical appendix which accompanied the original version of this manuscript is available upon request. This appendix provides a formal justification for (10) and (11) and utilizes Kenneth Arrow and Alain Enthoven's Theorem 1 to demonstrate that these conditions are sufficient to establish existence. A similar proof is also available in my 1973 paper.

$$(10)\quad \frac{\partial B_t}{\partial u_t} - \phi_t \le \beta\lambda_{t+1}\left(-\frac{\partial g_t}{\partial u_t}\right) + \beta\eta_{t+1}\frac{\partial D_t}{\partial u_t} - \beta s^x_{t+1}\frac{\partial \omega^x_t}{\partial u_t} - \beta s^u_{t+1}\frac{\partial \omega^u_t}{\partial u_t}$$

$$\le \left[\sum_{j=t+1}^{T-1}\beta^{j-t}X_j\frac{\partial x_j}{\partial x_{t+1}} + \beta^{T-t}\frac{\partial\psi}{\partial x_T}\frac{\partial x_T}{\partial x_{t+1}}\right]\left(-\frac{\partial g_t}{\partial u_t}\right)$$

$$+\left[\sum_{j=t+1}^{T-1}\beta^{j-t}Y_j\frac{\partial y_j}{\partial y_{t+1}} + \beta^{T-t}\frac{\partial\psi}{\partial y_T}\frac{\partial y_T}{\partial y_{t+1}}\right]\left(\frac{\partial D_t}{\partial u_t}\right)$$

$$-\left[\sum_{j=t+1}^{T-1}\beta^{j-t}\left\{\frac{\partial B_j}{\partial W^x_j} + \left[\sum_{i=j+1}^{T-1}\beta^{i-j}X_i\frac{\partial x_i}{\partial x_{j+1}} + \beta^{T-j}\frac{\partial\psi}{\partial x_T}\frac{\partial x_T}{\partial x_{j+1}}\right]\frac{\partial g_j}{\partial W^x_j}\right\}\right.$$

$$\left.\cdot\frac{\partial W^x_j}{\partial W^x_{t+1}} + \beta^{T-t}\frac{\partial\psi}{\partial W^x_T}\frac{\partial W^x_T}{\partial W^x_{t+1}}\right]\left(\frac{\partial\omega^x_t}{\partial u_t}\right) - \left[\sum_{j=t+1}^{T-1}\beta^{j-t}Z^u_j\frac{\partial W^u_j}{\partial W^u_{t+1}}\right.$$

$$\left.+\beta^{T-t}\frac{\partial\psi}{\partial W^u_T}\frac{\partial W^u_T}{\partial W^u_{t+1}}\right]\left(\frac{\partial\omega^u_t}{\partial u_t}\right)$$

$$(11)\quad \frac{\partial B_t}{\partial v_t} \le \beta\eta_{t+1}\frac{\partial D_t}{\partial v_t} - \beta s^v_{t+1}\frac{\partial\omega^v_t}{\partial v_t} = \left[\sum_{j=t+1}^{T-1}\beta^{j-t}Y_j\frac{\partial y_j}{\partial y_{t+1}} + \beta^{T-t}\frac{\partial\psi}{\partial y_T}\frac{\partial y_T}{\partial y_{t+1}}\right]\left(\frac{\partial D_t}{\partial v_t}\right)$$

$$-\left[\sum_{j=t+1}^{T-1}\beta^{j-t}\left\{\frac{\partial B_j}{\partial W^v_j} - \left[\sum_{j=t+1}^{T-1}\beta^{i-j}Y_i\frac{\partial y_i}{\partial y_{j+1}} + \beta^{T-j}\frac{\partial\psi}{\partial y_T}\frac{\partial y_T}{\partial y_{j+1}}\right]\frac{\partial D_j}{\partial W^v_j}\right\}\frac{\partial W^v_j}{\partial W^v_{t+1}}\right.$$

$$\left.+\beta^{T-t}\frac{\partial\psi}{\partial W^v_T}\frac{\partial W^v_T}{\partial W^v_{t+1}}\right]\left(\frac{\partial\omega^v_t}{\partial v_t}\right)$$

played as equations (10) and (11).[8] Note that

$$X_j = \partial B_j/\partial x_j + \phi_j\partial h_j/\partial x_j$$
$$Y_j = \partial B_j/\partial y_j + \phi_j\partial h_j/\partial y_j$$
$$Z^u{}_j = \partial B_j/\partial W^u_j + \phi_j\partial h_j/\partial W^u_j$$

If the strict inequality holds for (10) or (11), u_t or v_t is zero; however, if u_t or v_t is positive, relation (10) or (11) is a strict equality. In addition, if constraint (1), i.e., $u_t \le h(x_t, y_t, W^u_t)$ is binding in period t, the multiplier $\phi_t \ge 0$ for that period and if (1) is not binding $\phi_t = 0$.

As relationships (10) and (11) indicate, the decision process conforms to Markovian dependence structure; i.e., given x_t, y_t, $W^k_t(k=x, y, u)$, the optimal levels of u_t and v_t are independent of u_j and v_j for $j<t$. The stock or state variables completely summarize the influence of all previous decisions upon current and future optimal actions. To be sure, however, the optimal rates of production and investment must be determined simultaneously for all periods later than t in order to derive the optimal rate for period t. In view of this dependence structure in the optimal actions over time, there is no loss in generality if the initial optimal decisions, i.e., at $t=0$, are investigated for any T-period planning horizon problem with x_0, y_0, and W^k_0 given.

For the initial period, if we assume u_0 and v_0 are positive, then by (10) and (11), the

[8] Note that the subscript t on the functions h, g, D, and ω^i in the partial derivatives of these functions implies the partials are evaluated for solution values of their arguments in period t. For example $\partial h_t/\partial x_t$ is the partial derivative of $h(x, y, W^u)$ with respect to x, evaluated at $x=x_t$, $y=y_t$, and $W^u=W^u_t$.

$$(12)\quad \frac{\partial B_0}{\partial u_0} - \phi_0 = \left[\sum_{j=1}^{T-1} \beta^j X_j \frac{\partial x_j}{\partial x_1} + \beta^T \frac{\partial \psi}{\partial x_T}\frac{\partial x_T}{\partial x_1}\right]\left(-\frac{\partial g_0}{\partial u_0}\right)$$
$$+ \left[\sum_{j=1}^{T-1} \beta^j Y_j \frac{\partial y_j}{\partial y_1} + \beta^T \frac{\partial \psi}{\partial y_T}\frac{\partial y_T}{\partial y_1}\right]\left(\frac{\partial D_0}{\partial u_0}\right)$$
$$- \left[\sum_{j=1}^{T-1} \beta^j \left\{\frac{\partial B_j}{\partial W_j^x} + \left[\sum_{i=j+1}^{T-1} \beta^{i-j} X_i \frac{\partial x_i}{\partial x_{j+1}} + \beta^{T-j}\frac{\partial \psi}{\partial x_T}\frac{\partial x_T}{\partial x_{j+1}}\right]\frac{\partial g_j}{\partial W_j^x}\right\}\frac{\partial W_j^x}{\partial W_1^x}\right.$$
$$\left. + \beta^T \frac{\partial \psi}{\partial W_T}\frac{\partial W_T}{\partial W_1^x}\right]\left(\frac{\partial \omega_0^x}{\partial u_0}\right) - \left[\sum_{j=1}^{T-1} \beta^j Z_j^u \frac{\partial W_j^u}{\partial W_1^u} + \beta^T \frac{\partial \psi}{\partial W_T^u}\frac{\partial W_T^u}{\partial W_1^u}\right]\left(\frac{\partial \omega_0^u}{\partial u_0}\right)$$
$$= \beta\lambda_1\left(-\frac{\partial g_0}{\partial u_0}\right) + \beta\eta_1\frac{\partial D_0}{\partial u_0} - \beta s_1^x \frac{\partial \omega_0^x}{\partial u_0} - \beta s_1^u \frac{\partial \omega_0^u}{\partial u_0}$$

$$(13)\quad \frac{\partial B_0}{\partial v_0} = \left[\sum_{j=1}^{T-1} \beta^j Y_j \frac{\partial y_j}{\partial y_1} + \beta^T \frac{\partial \psi}{\partial y_T}\frac{\partial y_T}{\partial y_1}\right]\left(\frac{\partial D_0}{\partial v_0}\right)$$
$$- \left[\sum_{j=1}^{T-1} \beta^j \left\{\frac{\partial B_j}{\partial W_j^v} - \left[\sum_{i=j+1}^{T-1} \beta^{i-j} Y_j \frac{\partial y_i}{\partial y_{i+1}} + \beta^{T-j}\frac{\partial \psi}{\partial y_T}\frac{\partial y_T}{\partial y_{j+1}}\right]\frac{\partial D_j}{\partial W_j^v}\right\}\frac{\partial W_j^v}{\partial W_1^v}\right.$$
$$\left. + \beta^T \frac{\partial \psi}{\partial W_T^v}\frac{\partial W_T^v}{\partial W_1^v}\right]\left(\frac{\partial \omega_0^v}{\partial v_0}\right) = \beta\eta_1\frac{\partial D_0}{\partial v_0} - \beta s_1^v \frac{\partial \omega_0^v}{\partial v_0}$$

optimum rates of production and investment are obtained as equations (12) and (13). These two equations will serve as the basis for the economic interpretations of the major differences between the present model and models previously advanced, and the equilibrium conditions which are treated in the following section.

II. Economic Interpretations and Equilibrium Conditions

A number of economic interpretations can be offered for the conditions (10) through (13). We shall discuss two of these by first examining the characteristics of the optimal rates of utilization u_t and investment v_t separately and then some implications of the model which recognize that the solutions for u_t and v_t must be determined simultaneously. In each of these instances we distinguish the results obtained here from those found in Burt and Cummings. This discussion is followed by investigation of the equilibrium conditions for the natural resource, capital, and knowledge stocks assuming an infinite planning horizon.

The optimum time path for natural resource utilization at $t=0$ is represented by (12). As in the case of the Burt and Cummings model, the left-hand side of (12) represents the marginal social benefits attributable to an increment in resource use at $t=0$ less the imputed value of the constraint (1). The latter value, if positive, is the initial marginal cost in sacrificed social benefits associated with the binding constraint $u_0=h(x_0, y_0, W_0^u)$. If this constraint is not binding, then this value is, of course, zero (i.e., $\phi_0=0$). This value whether zero or positive will be referred to subsequently as the "boundary cost."

On the right-hand side of (12), the partial derivatives $\partial g_0/\partial u_0$, $\partial D_0/\partial u_0$, and $\partial \omega_0^k/\partial u_0$ ($k=x, u$) represent the marginal influence of u_0 on resource, capital, and knowledge stocks. Recall that $\partial D_0/\partial u_0 \geq 0$, $\partial \omega_0^k/\partial u_0 \geq 0$,

and $\partial g_0/\partial u_0 \geq 0$ where $\partial D_0/\partial u_0$ is the marginal capital consumption from current utilization, $\partial \omega_0^k/\partial u_0$ is the marginal learning for process k ($k=x, u$) resulting from an increment in resource use at t=0, and for most applications in which resource use is measured in the same units as resource stocks $\partial g_0/\partial u_0 = -1$. It should be noted that the first two terms are equivalent to the right-hand side derived from the Burt and Cummings model. Hence the principal difference between the present model and their model is the inclusion of two additional terms on the right-hand side of (12) which result from the learning by doing specifications (4) and (6).

The factors $\beta\lambda_1$, $\beta\eta_1$, and $\beta s_1^k (k=x, u)$ appearing on the right-hand side are easily interpreted. The factor $\beta\lambda_1$ is the discounted value of increments to x for all future periods associated with an increment to resource stocks in period one; $\beta\eta_1$ is the discounted value of increments to y over future periods associated with an increment to capital stock in period one; βs_1^x is the discounted value of increments to W^x for future periods resulting from an increment to resource stock knowledge in period one; and βs_1^u is the discounted value of increments to W^u in future periods associated with an increment to resource production knowledge in period one. When these values are multiplied by their associated partial derivatives discussed in the previous paragraph, the first term on the right-hand side of (12) may be interpreted as the discounted marginal value of a unit of resource maintained in stocks rather than used in current production; the second as the discounted marginal value of capital consumed by an increment in current production; the third as the discounted marginal value of resource stocks learning resulting from an increment in current production; and similarly the fourth term may be interpreted as the discounted marginal value of resource production learning emanating from an increment in current utilization.

The sum of the first two terms on the right-hand side of (12) may be interpreted simply as "user cost." This sum reflects, as Burt and Cummings quoting Anthony Scott point out, "... the present value of future profit (social benefits) foregone by a decision to produce a unit of output today (at t=0)" (p. 583). Likewise, the third and fourth terms represent a marginal value of learning or additional knowledge reflecting the present value of future social benefits resulting from a decision to produce a unit of output today. Hence (12) suggests that the proper decision rule is to increase production at t=0 until net marginal social benefits (marginal social benefits less boundary cost) equal user costs less the present value of future marginal benefits resulting from learning at t=0. A similar interpretation for this decision rule would hold if learning was measured on the cost side, i.e., increase production at t=0 until net marginal social benefits at t=0 equal user costs less the present value of the marginal reduction in future costs resulting from learning at t=0.

The above economic interpretation may be further clarified by examining the separate components of each term on the right-hand side of (12). Since detailed explanations of the components entering the first two terms are considered by Burt and Cummings, we shall be concerned here only with the third and fourth terms. Specifically, the separate components of βs_1^u are analyzed in some detail, and an almost identical explanation is offered for βs_1^x.

The term $\phi_j(\partial h_j/\partial W_j^u)$ is the implicit or shadow value of a marginal unit of resource production knowledge available in period j to relax the constraint $u_j \leq h(x_j, y_j, W_j^u)$; and the sum $\partial B_j/\partial W_j^u + \phi_j(\partial h_j/\partial W_j^u) = Z_j^u$ is the marginal value of resource production knowledge during period j. The first term is the direct effect of learning rewards in period j and the second term the indirect effect of learning rewards through constraint (1). In other words, *ceteris paribus*, this sum is the increment to benefits in period j associated with an increment in resource production knowledge during the same period. Obviously, in the terminal period t=T, this same measure is $\partial\psi/\partial W_t^u$. All of these marginal values when multiplied by $\partial W_j^u/\partial W_1^u$ are values which are transferred to a com-

parable increment in resource production knowledge in period one rather than period j. The specific components of the third term (i.e., βs_1^x) although more complex, may be similarly diagnosed. The additional complexity associated with this term results from the dependence of s_1^x on the implicit value associated with the resource stock constraint (2) over the periods $t=2$ through $t=T-1$, i.e., $\lambda_2, \ldots, \lambda_{T-1}$. Nevertheless, the sum $\partial B_j/\partial W_j^x + \beta\lambda_{j+1}(\partial g_j/\partial W_j^x) = Z_j^x$ may be regarded as the marginal value of resource stock knowledge during period j and when multiplied by $\partial W_j^x/\partial W_1^x$, this value is comparable to an increment to such knowledge stocks in period one instead of period j.

A second possible interpretation, perhaps more appealing than the above explanation, may be seen by rewriting (12) as

$$\text{(12a)}\quad \beta\lambda_1\left(-\frac{\partial g_0}{\partial u_0}\right)+\beta\eta_1\frac{\partial D_0}{\partial u_0}-\left(\frac{\partial B_0}{\partial u_0}-\phi_0\right) = \beta s_1^x\left(\frac{\partial \omega_0^x}{\partial u_0}\right)+\beta s_1^u\left(\frac{\partial \omega_0^u}{\partial u_0}\right)$$

The left-hand side of this expression, given previous assumptions, is the divergence in the initial period from what would otherwise be the maximum benefits with respect to resource use if learning by experience were disregarded in the decision-making process. Hence, it represents the marginal cost of additional knowledge in terms of current benefits foregone or simply the marginal cost of knowledge accumulation. On the other hand, what now appears on the right-hand side is the discounted marginal value of resource learning since (i) as previously noted βs_1^x, βs_1^u represent the discounted increments in social benefits for all future periods resulting from an addition to knowledge (resource stock and use) in period one; and (ii) $\partial\omega_0^x/\partial u_0$, $\partial\omega_0^u/\partial u_0$ represent the marginal learning obtained for resource stocks and use, respectively, by utilizing an optimal amount of u_0.

The above interpretation implies that condition (12) requires imposing a wedge between costs and social benefits such that current net marginal social benefits are less than current user costs. It also suggests that the usual natural resource models which neglect learning result in an understatement of the optimal level of current resource use. If we characterize the Burt-Cummings formulation as the usual case, then the public agency envisaged in the present model "overexpands" by comparison with the usual case. This interpretation demonstrates in addition that even though resource learning is costless (in the sense that it is an automatic consequence of resource use), rational decision making requires incurring a positive opportunity cost (left-hand side of (12a)) to gain more knowledge than would be obtained by maximizing the present value of social benefits neglecting the learning by experience equations (4) and (6). Hence, the specification admitting endogenous technological change of the learning by doing type recognizes that public decision makers may find it beneficial to incur opportunity costs in the current period so as to, in effect, substitute knowledge for resource and capital investments in some later period.

To investigate the optimal rate of investment our attention turns to equation (13). The two interpretations offered for condition (12) clearly hold as well for equation (13). The first is most easily treated after multiplying condition (13) by (-1). The left-hand side of the resulting expression is the initial marginal social cost of capital investment, while the right-hand side is the sum of (i) the initial discounted marginal value of capital stocks in all future periods associated with an increment to current investment, and (ii) the initial discounted marginal value of capital knowledge in all future periods associated with an increment to current investment. The economic interpretation of the first term (i) on the right-hand side of (13) is equivalent to the second user cost term in (12) except that the weight is $-\partial D_0/\partial v_0$ (the marginal effect on initial depreciation of an increment in gross investment in the initial period) rather than $\partial D_0/\partial u_0$, while the economic interpretation of the second term (ii) is equivalent to the third or fourth terms on the right-hand side of (12) except here the relevant learning process is associated with gross investments

and hence capital stocks rather than resource stocks or rates of utilization.

The second or opportunity cost interpretation for learning emanating from capital investments may be indicated by restating (13) as

$$\text{(13a)}\quad \left(-\frac{\partial B_0}{\partial v_0}\right) - \beta\eta_1\left(-\frac{\partial D_0}{\partial v_0}\right) = \beta \overset{y}{s_1}\frac{\partial \overset{y}{\omega_0}}{\partial v_0}$$

As in the case of (12a) this expression suggests that the marginal cost of additional capital knowledge in terms of initial benefits foregone (left-hand side) be equated to the initial discounted marginal value of learning associated with the investment process.

The above discussion treats the characteristics of the optimal production and investment rates separately even though they must be determined simultaneously within the specified model. To provide some flavor for the interdependencies among the solutions for u_t and v_t, we may solve (13) for $\beta\eta_1$ and substitute the result into the second term of (12) to obtain

$$\text{(14)}\quad \frac{\partial B_0}{\partial u_0} - \phi_0 = \beta\lambda_1\left(-\frac{\partial g_0}{\partial u_0}\right) - \frac{\partial B_0}{\partial v_0}\frac{\partial v_0}{\partial u_0} - \beta\overset{x}{s_1}\frac{\partial \overset{x}{\omega_0}}{\partial u_0} - \beta\overset{y}{s_1}\frac{\partial \overset{y}{\omega_0}}{\partial v_0}\frac{\partial v_0}{\partial u_0} - \beta\overset{u}{s_1}\frac{\partial \overset{y}{\omega_0}}{\partial u_0}$$

On the right-hand side of (14), the first, third, and fifth terms are equivalent to the first, third, and fourth terms, respectively, on the right-hand side of (12). The second term in (12) becomes $-(\partial B_0/\partial v_0+\beta \overset{y}{s_1}\partial\overset{y}{\omega_0}/\partial v_0)\partial v_0/\partial u_0$ which measures the capital and knowledge capital costs associated with an increment in current production. Clearly, if capital stock is held constant, adjustments in capital at the margin must occur by varying v such that the net effect on depreciation will be zero. Hence, $-(\partial B_0/\partial v_0+\beta \overset{y}{s_1}\partial\overset{y}{\omega_0}/\partial v_0)$ may be regarded as the marginal social cost associated with current investment, and multiplication by $\partial v_0/\partial u_0$ converts it to marginal social cost associated with current resource production. This conversion is based on the depreciation function $D(\cdot)$ and thus is related to capital consumption associated with current production.

The opportunity cost interpretation for knowledge accumulation advanced in context of (12) and (13) becomes for (14)

$$\beta\lambda_1\left(-\frac{\partial g_0}{\partial u_0}\right) + \frac{\partial B_0}{\partial v_0}\left(-\frac{\partial v_0}{\partial u_0}\right) - \left(\frac{\partial B_0}{\partial u_0} - \phi_0\right) = \beta\left(\overset{x}{s_1}\frac{\partial \overset{x}{\omega_0}}{\partial u_0} + \overset{y}{s_1}\frac{\partial \overset{y}{\omega_0}}{\partial v_0}\frac{\partial v_0}{\partial u_0} + \overset{u}{s_1}\frac{\partial \overset{u}{\omega_0}}{\partial u_0}\right)$$

In other words, marginal social costs less net marginal social benefits (which may be interpreted as the marginal social cost of additional resource knowledge), both with respect to initial resource use, is equated to present discounted marginal value of learning which now includes the marginal values of resource stock learning, of capital stock learning, and of resource utilization learning, all with respect to initial resource use.

As previously indicated, the equilibrium conditions for the state variables generated by equations (2) through (6) will be examined within the context of an infinite planning horizon.[9] Given the existence of an equilibrium, the state variables all assume constant values for successive periods of time which in turn suggests that

$$\text{(15)}\quad g(\ )=D(\ )=\omega^x(\ )=\omega^y(\ )=\omega^u(\ )=0$$

for all t. For (15) to hold, clearly $u_t=u_{t+1}=$

[9] To establish the existence and uniqueness of the criterion function (9) under an infinite planning horizon, i.e., that the Max $B(\ )$ has uniform convergence as $T\rightarrow\infty$, we shall impose the same restrictive assumptions employed by Burt and Cummings. These assumptions allow us to appeal to a general theorem of Richard Bellman (p. 121) on the functional equation of dynamic programming. They are: (a) $B(\ )$ is not explicitly a function of time; (b) x, y, W^x, W^y, W^u are confined to a bounded region, say R; (c) the difference equations (2) through (6) are consistent with $(x, y, W^x, W^y, W^u)\in R$; and (d) the function $B(\ )$ is uniformly bounded for all values of its arguments belonging to the constraint set specified by (1) through (8) and $(x, y, W^x, W^y, W^y)\in R$. It should be noted, however, that assumption (a) is not nearly as restrictive in the present model as in the Burt and Cummings model since here $B(\ )$ is treated as an explicit function of the endogenous technical knowledge variables W^x, W^y, and W^u. Under assumptions (a) through (d) as well as earlier assumptions related to continuity and quasi concavity, either an equilibrium must exist, or resource stocks become zero in a finite period.

. . . and $v_t = v_{t+1} =$ . . . . The implications of these variable constancies may be examined with respect to the necessary conditions represented by equations (12) and (13).

Since the functions entering (12) and (13) for an equilibrium state are constant over time, the temporal subscript (t) will be discarded and all arguments are fixed at constant values u, v, x, y, W^x, W^y, and W^u. The Lagrange multipliers λ_t, η_t, ϕ_t, s_t^x, s_t^p, and s_t^u will be constant in equilibrium and thus the t subscript will also be deleted on these variables. Given this background, equations (12) and (13) will be simplified under the assumption that the summations appearing on the right-hand sides of these expressions are convergent as $T\rightarrow\infty$. By (15), the partial derivatives of the state difference equations (2) through (6) with respect to first period levels of the state variables may be simplified[10] and when substituted into (12) upon letting $T\rightarrow\infty$, we have

$$(16)\quad \frac{\partial B}{\partial u} - \phi = \beta X(1-\beta q^x)^{-1}\left(-\frac{\partial g}{\partial u}\right) + \beta Y(1-\beta q^y)^{-1}\left(\frac{\partial D}{\partial u}\right) - \beta\left[\frac{\partial B}{\partial W^x} + \beta X(1-\beta q^x)^{-1}\frac{\partial g}{\partial W^x}\right] \cdot (1-\beta q^{\omega x})^{-1}\left(\frac{\partial \omega^X}{\partial u}\right) - \beta Z^u(1-\beta q^{\omega u})^{-1}\frac{\partial \omega^u}{\partial u}$$

where $q^x = (1+\partial g/\partial x)/(1+r)$
$q^y = (1-\partial D/\partial y)/(1+r)$
$q^{\omega x} = (1+\partial\omega^x/\partial W^x)/(1+r)$
$q^{\omega u} = (1+\partial\omega^u/\partial W^u)/(1+r)$

Proceeding in a similar fashion for (13), we obtain the corresponding result[11]

$$(17)\quad \frac{\partial B}{\partial v} = \beta Y(1-\beta q^y)^{-1}\left(\frac{\partial D}{\partial v}\right) - \beta\left[\frac{\partial B}{\partial W^y} - \beta Y(1-\beta q^y)^{-1}\frac{\partial D}{\partial W^y}\right] \cdot (1-\beta q^{\omega y})^{-1}\left(\frac{\partial \omega^y}{\partial v}\right)$$

where $q^{\omega y} = (1+\partial\omega^y/\partial W^y)/(1+r)$

As for (12) and (13), the same economic interpretations may be offered for (16) and (17). For example, the opportunity cost of knowledge interpretation may be isolated by multiplying equation (16) by (-1) and moving the first two terms on the right-hand side to the left-hand side. On the left-hand side we would then have user cost (composed of two terms: the "capitalized value" of an addition to resource stocks multiplied by the marginal effect current utilization has on next period's resource stocks, and the capitalized value of an addition to capital stocks multiplied by the marginal effect current utilization has on next period's capital stocks) less net marginal social benefits with respect to current resource use. The right-hand side involves the sum of two terms associated with the capitalized marginal value of resource learnings, viz., 1) the capitalized value of an addition to resource knowledge stocks multiplied by the marginal effect current utilization has on next period's level of these stocks, and 2) the capitalized value of an addition to resource production knowledge multiplied by the marginal effect current utilization has on next period's resource production knowledge. For each of these interpretations, capitalized value is employed to denote the reduction of a perpetual flow to a present value measure when the "discounting factor" is q^x, q^y, $q^{\omega x}$, or $q^{\omega u}$. These discounting factors represent the combined influence of an economic factor $\beta = 1/(1+r)$, and the stock compounding

[10] That is

$$\partial x_j/\partial x_1 = (1+\partial g/\partial x)^{j-1}$$
$$\partial y_j/\partial y_1 = (1-\partial D/\partial y)^{j-1}$$
$$\partial W_j^x/\partial W_1^x = (1+\partial\omega^x/\partial W^x)^{j-1}$$
$$\partial W_j^y/\partial W_1^y = (1+\partial\omega^y/\partial W^y)^{j-1}$$

and $$\partial W_j^u/\partial W_1^y = (1+\partial\omega^u/\partial W^u)^{j-1}$$

[11] The two statements (16) and (17) presume that the infinite sums evolving in (12) and (13) as $T\rightarrow\infty$ are convergent. This convergence obtains if $|q^k|<1$, $k = x, y, \omega x, \omega y, \omega u$. Assumptions (b) and (c) of fn. 9 indirectly insure that these conditions are satisfied. Note also that the conditions on q^x and q^y are equivalent to those stated in Burt and Cummings (p. 585) and similar to those obtained by James Quirk and Smith (p. 16).

factor $1+\partial g/\partial x$, or the stock discounting factors $1-\partial D/\partial y$, $1+\partial\omega^x/\partial W^x$, and $1+\partial\omega^u/\partial W^u$. Obviously, an analogous interpretation could be advanced for (17).

The equilibrium counterpart of (14) can be obtained by solving (17) for $\beta\eta$ and substituting the result in (16), i.e.,

$$(18)\quad \frac{\partial B}{\partial u} - \phi = \beta X(1-q^x)^{-1}\left(-\frac{\partial g}{\partial u}\right) - \frac{\partial B}{\partial v}\frac{\partial v}{\partial u} - \beta\left[\frac{\partial B}{\partial W^v} - \beta Y(1-q^v)^{-1}\frac{\partial D}{\partial W^v}\right] \cdot (1-q^{\omega v})^{-1}\frac{\partial\omega^v}{\partial v}\frac{\partial v}{\partial u} - \beta\left[\frac{\partial B}{\partial W^x} + \beta X(1-q^x)^{-1}\frac{\partial g}{\partial W^x}\right] \cdot (1-q^{\omega x})^{-1}\left(\frac{\partial\omega^x}{\partial u}\right) - \beta Z^u(1-q^{\omega u})^{-1}\left(\frac{\partial\omega^u}{\partial u}\right)$$

Again, two alternative economic interpretations might be offered for the result (18). We simply note here that the second term on the right of (18) is a measure of the implicit cost while the third term is a measure of the implicit learning value in capital consumption associated with an addition to current resource use. This value (cost) is expressed as the value (cost) of current investment required to maintain capital knowledge (capital) stocks. In (16) the second and third terms of (18) are simply expressed as the implicit cost of incrementally less capital stock over future periods. The latter implicit cost as (18) demonstrates is actually a net cost measure, i.e., net of the marginal value of learning associated with the capital investment process.

An equilibrium solution for the state and decision variables may be obtained from solving a system of seven equations composed of (15), (16), and (17), assuming the constraint $u \leq h(x, y, W^u)$ is not binding in the equilibrium state. Solving this system would provide equilibrium solutions for x, y, W^x, W^v, W^u, u, and v. If the constraint on resource use in (1) is binding, of course, the equation $u=h(x, y, W^u)$ must be added to the above system, and boundary costs ϕ is obtained as an eighth variable of equilibrium. For each of these two cases, a reasonably simple analysis is involved to determine under an optimal policy the equilibrium resource, capital, and knowledge stocks as well as the equilibrium rates of resource use and gross capital investment.

REFERENCES

J. K. Arrow, "The Economic Implications of Learning by Doing," *Rev. Econ. Stud.*, June 1962, *29*, 155–73.

——— and A. C. Enthoven, "Quasi-Concave Programming," *Econometrica*, Oct. 1961, *4*, 779–800.

R. E. Bellman, *Dynamic Programming*, Princeton 1957.

O. Burt and R. Cummings, "Production and Investment in Natural Resource Industries," *Amer. Econ. Rev.*, Sept. 1970, *60*, 576–90.

R. Cummings and O. Burt, "The Economics of Production from Natural Resources: Note," *Amer. Econ. Rev.*, Dec. 1969, *59*, 985–90.

H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," in J. Neyman, ed., *Proceedings of Second Berkeley Symposium On Mathematical Statistics and Probability*, Berkeley 1951, 481–92.

J. V. Krutilla, "Conservation Reconsidered," *Amer. Econ. Rev.*, Sept. 1967, *57*, 777–86.

J. Quirk and V. Smith, "Dynamic Economic Models of Fishing," in A. Scott, ed., *Proceedings of the H. R. MacMillan Fisheries Economics Symposium*, Vancouver 1969.

G. C. Rausser, C. Willis, and P. Frick, "Learning, External Benefits, and Subsidies in Water Desalination," *Water Resources Res.*, Dec. 1972, *8*, 1385–1400.

———, "Learning, Production, and Investment in Natural Resource Industries," Rept. 7336, Univ. Chicago Center Math. Stud. Bus. Econ., 1973.

A. Scott, "The Theory of the Mine under Conditions of Certainty," in M. Gaffney, ed., *Extraction Resources and Taxation*, Madison 1967.

V. Smith, "Economics of Production from Natural Resources," *Amer. Econ. Rev.*, June 1968, *58*, 409–31.

Cropsharing Tenancy in Agriculture: Comment

By D. M. G. Newbery*

In a recent contribution to this *Review*, P. K. Bardhan and T. N. Srinivasan (hereafter B-S) constructed a simple theoretical model of sharecropping from which they were able to make several predictions, and which they tested against Indian data. They claim to have constructed a competitive theory of share tenancy in which the equilibrium share rental r^* is parametrically given to all agents, which they contrast with an earlier attempt at a general equilibrium model of sharecropping by S. N. S. Cheung (1968). They criticize Cheung for ignoring his own assumptions " . . . that the percentage share and area rented under share tenancy . . . are competitively determined in the market" (1968, p. 1120) when analyzing share tenancy contracts, and B-S argue that the difference in their conclusions " . . . lies in the kind of maximization process Cheung carried out in Section III, pp. 1113–14 of his paper. There he maximizes only from the landlord's point of view, whereas in this paper we determine the demand side from maximization by the tenant, just as the supply side is determined from landlord's maximizing decision" (p. 52).

This note argues that Cheung was correct in describing a competitive theory of share tenancy, though the incompleteness of his treatment appears to have misled B-S and possibly others. It follows that the analysis of B-S is incorrect, and that any attempt to modify their model while retaining the spirit of the original will ironically give rise to a pure monopoly equilibrium. Any landlord will be faced with excess demand for land and will be able to set whatever share-rental he chooses, regardless of the decisions of all other landlords. It can also be shown that the theoretical predictions are invalidated, and that under certain assumptions the monopoly equilibrium will be unstable, tending to be replaced by the equilibrium described by Cheung.

I begin by explaining the assumptions underlying the models since these are implicit and ambiguous, if not internally inconsistent, in the B-S model. Next I show that Cheung's model is correct in its own terms, and moreover its results continue to hold in a more general model which explicitly recognizes uncertainty. A brief resumé of the B-S model will show the weakness of that model, and prepare for the derivation of the remaining results.

A close study of the explicit and implicit assumptions of the B-S model leaves a number of unanswered questions, so that the model is ambiguous in a number of important respects. The following assumptions are reasonably unambiguous:

1) Land and tenant labor are homogeneous.
2) Perfectly certain wage labor and employment is always available at an exogenously determined and parametrically given wage rate w.
3) Landlords are numerous and noncollusive.

With a perfect labor market (assumption 2), nothing is gained by making agents maximize the utility of consumption and leisure compared with the simpler assumption of maximizing income net of the imputed cost of labor, so that elaboration of B-S is unnecessary.

The remaining assumptions, while not made explicitly by B-S, form a useful starting point for further analysis.

4) Environmental risk (arising from weather, pests, disease, etc.) is multiplicative so that output from land H and labor L is $\theta F(H, L)$ where θ is a random variable with mean unity. In the riskless case, $\theta = 1$.
5) Agents know the distribution of θ and

* Lecturer in economics, Cambridge University and Fellow of Churchill College, Cambridge, England.

maximize expected utility of income $EU(Y)$.

6) The production function exhibits constant returns to scale, so that $F(H, L) = Hf(l) = Lg(h)$, with $h = l^{-1} = H/L$.
7) Agents can mix contracts, and there are no transaction or information costs.

The notation has been slightly simplified and additional assumptions which are implicit in most treatments of perfect competition have been made explicit. Assumptions 4) and 5) allow us to treat risk rather more generally than the approach suggested by B-S in their Section VI. Assumption 6) is not essential for the general argument but allows considerable simplification. If anything, diminishing returns will be less favorable to the existence of sharecropping; conversely, increasing returns (or U-shaped cost curves) are more favorable—a point argued elsewhere by the author (1973b). In fact, B-S at one stage seem to make the constant returns assumption since they allow tenants to hire parcels of size h from each of n landlords and derive output $nF(h, L_1)$ instead of $F(nh, nL_1)$, which would be required by strict concavity (see their equations (20) and (21), p. 53).

I. A Competitive Theory of Share Tenancy

Consider first the simple model suggested by Cheung which ignores uncertainty and allows tenants a choice between wage and rental contracts (or a mixture of them) and allows landlords to specify minimum levels of tenant labor input on tenancies. In equilibrium all agents must be maximizing their utility (or profit net of imputed costs), taking all other contracts and prices as given. If the combination of contracts is to yield a simple sum, we must have constant returns as has been noted, in which case tenant net income Y is

$$(1)\quad Y = \sum_i H_i(1 - r_i)f(l_i) - w \sum_i H_i l_i$$

Here the ith contract allows the tenant to hire H_i land at a share rent r_i on condition that the tenant apply at least l_i^* units of labor per acre. The tenant maximizes Y subject to this constraint:

$$(2)\quad \left.\begin{aligned}\frac{\partial Y}{\partial l_i} = H_i[(1-r_i)f'(l_i) - w] \leqq 0 \\ l_i \geqq l_i^*\end{aligned}\right\} \text{complementarily}$$

Demand for the ith contract will be nonzero if

$$(3)\quad \frac{\partial Y}{\partial H_i} = (1 - r_i)f(l_i) - wl_i \geqq 0$$

If this is satisfied with equality, $(1-r_i)f'(l_i) < w$ since f is a concave function, therefore $l_i = l_i^*$, and equation (3) must hold for $l_i = l_i^*$.

The landlord now chooses r_i, l_i^* to maximize his income per acre from rented land subject to the contract remaining attractive, i.e., subject to equation (3).

$$(4)\quad \underset{r,l^*}{\text{Max}}\ rf(l^*) + \lambda[(1 - r)f(l^*) - wl^*]$$

(dropping subscripts) whence

$$(5)\quad \lambda = 1, \qquad (1 - r)f(l^*) = wl^*$$

$$(6)\quad f'(l^*) = w$$

Equation (5) was Cheung's condition of equilibrium which B-S objected to in footnote 9, p. 52, where they argued that *tenant* optimization required $(1-r)F_2 = w$. This ignores the contractual constraint explicitly introduced by Cheung (and extensively illustrated with examples of contracts (1969, pp. 76–77)), which equation (2) recognizes. Landlord optimization leads to efficiency; in equation (6) the marginal product of labor is forced to equality with the wage rate and also leads to the equilibrium condition. In this equilibrium the rental share r is equated everywhere, workers are indifferent between working and becoming tenants so there is no excess supply of tenants, nor excess demand for land by tenants. It is clear that the ability of the landlord, in this certain world, to specify intensity and share rent gives him the same power to extract the rental surplus as a fixed rent contract, subject to the same constraint that the contract remain attractive, so that the argument carries over to the case of diminishing returns.

Bardhan and Srinivasan further observe that Cheung does not give the landlords the option to cultivate, so that the range of

choices available to the economic agents is not as large as in their treatment. A full treatment ought to also include the *possibility* of fixed rent tenure and explicitly recognize uncertainty: I now turn to this generalization. With constant returns tenants maximize $EU(Y)$ subject to a minimum contractual intensity l^*:

$$(7) \qquad Y = (1-r)\theta H_1 f(l) - wlH_1 + \theta F(H, L) - wL - RH$$

$$(8) \qquad \frac{\partial EU(Y)}{\partial L} = F_2 EU'\theta - wEU' = 0$$

$$(9) \qquad \frac{\partial EU(Y)}{\partial H} = F_1 EU'\theta - REU' = 0$$

For share tenancies to be demanded,

$$(10) \qquad \frac{\partial EU(Y)}{\partial H_1} = (1-r)f(l^*)EU'\theta - wlEU' \geqq 0$$

This will be satisfied if, using equation (8),

$$(11) \qquad (1-r)f(l^*) \geqq l^* F_2$$

and the argument of equations (4) to (6) applies to reveal that the marginal product of labor is equated on all tenant-operated land. Moreover, landlords maximize $EU(W)$, say where, assuming that they have access to the same production function as tenants,

$$(12) \quad W = \theta r H_1 f(l^*) + \theta F(H, L) - wL + R(\overline{H} - H_1 - H)$$

whence, differentiating $EU(W)$ with respect to H, L, H_1:

$$(13) \qquad \frac{F_2}{w} = \frac{EU'}{EU'\theta} = \frac{F_1}{R} = \frac{rf(l^*)}{R}$$

This is the same set of conditions as equations (8) to (11), showing that positive levels of all contracts may be observed, and that production will be efficient everywhere. It is also easy to show that in this model share tenancies offer no advantage to either party over a mixture of $1-r$ units of labor applied to a fixed rent contract and r units to wage labor (see the author (1973a)).

Thus under constant returns and the other assumptions, a sharecropping equilibrium is identical to (and offers no advantages over) a fixed rent-wage labor equilibrium and can therefore be legitimately described as a competitive general equilibrium. This generalization of Cheung's result provides a useful reference point since it is so easily characterized as being an efficient equilibrium.

We can make a number of comments on the implicit assumptions of B-S. They assume that landlords have access to a different (and inferior) production function $G(H, L)$. This could be defended by supposing that hired labor needs supervision, the difference between the landlords' G and tenants' (own cultivators') F constituting the supervision or labor contract enforcement costs. This is somewhat at variance with their assumption that landlords' labor is as valuable as hired labor, but not seriously so. Next, fixed rent contracts give rise to a different (and inferior) production function than share rent contracts (see fn. 8, pp. 51–52). C. H. Hanumantha Rao has made the perceptive observation that the ease of contractual enforcement will depend, among other things, on the technical properties of the crop; specifically on the scope for entrepreneurship. For some crops it is clear what actions the cultivator ought to take and whether he has in fact taken them; these crops lend themselves to share tenancies. For others it is less clear, and a fixed rent contract will be favored. A share tenancy will therefore frequently specify the crop to be cultivated, so it might be argued that different contracts lead to different production functions. It takes more (and subtler) arguments to claim that the fixed rent production function will be inferior, since the conventional (and plausible) view is that share tenancies are more costly to enforce; fixed rent tenants can be left free to choose their crop and can presumably choose the share rent crop if this is attractive. A satisfactory explanation will require explicit uncertainty *and* further modifications to the simple model analyzed above. (Some examples are given in my 1973b paper, and the difficulties, which do not end there, are discussed further in my 1973a paper.)

II. Monopolistic Equilibrium in the Bardhan-Srinivasan Model

In contrast to Cheung's theory, share tenants in the B-S model are not constrained to apply a minimum intensity of cultivation, and can be represented as maximizing

$$(14) \qquad Y = (1 - r)F(H, L) - wL$$

so

$$(15) \qquad (1 - r)F_2 - w = 0$$

$$(16) \qquad \left.\begin{array}{r} (1 - r)F_1 \geqq 0 \\ H \leqq \overline{H} \end{array}\right\} \text{ complementarily}$$

which, under constant returns, is equivalent to:

$$(17) \qquad \left.\begin{array}{r} (1-r)F(H, L)-wL\geqq 0 \\ H\leqq\overline{H} \end{array}\right\} \text{ complementarily}$$

Bardhan and Srinivasan do not explicitly introduce the complementary slackness condition and so appear to be ignoring the possibility that the land available might be insufficient. At one point they are nearly forced to recognize that there may be excess demand for land; on page 53 they allow the landlord to divide up his holding into parcels and let tenants hold several parcels. They find in their equation (21) that tenants will have an excess demand for parcels, and since they further argue that there is for some unspecified reason a minimum parcel size below which a landlord will not choose to divide his holding, this excess demand is equivalent to an excess demand for land area. Ignoring this implication they press on to find a solution with the marginal product of land zero—this applied to a country noted for its overpopulation and land scarcity! It is now immediate that no competitive equilibrium can exist. For if F is strictly concave and $F_1 = 0$ then

$$(18) \qquad (1 - r)F(H, L) > wL$$

by Euler's theorem. In this case there will be an excess supply of tenants, since there is an elastic supply of labor implied by the exogenously given wage rate. Under either constant or diminishing returns it is far from obvious that the simultaneous solution of these equations (and corresponding equations for the landlords) will lead to a value for the amount of land leased which is feasible, given the finite amount available. Indeed, for many production functions such as the Cobb-Douglas the demand for land at zero marginal cost would be infinite. The problem is that the B-S equilibrium requires a scarce resource (land) to be priced effectively at zero with no excess demand—a requirement which in general cannot be met. If on the other hand land has to be rationed to force its marginal product above zero, it is clear that each landlord can set whatever rental r he likes. We can illustrate the consequences of this and simultaneously deal with the constant returns case in the following model. First we note that equation (17) holds only under constant returns, and only then can the supply of potential tenants equal demand given equations (15) and (16). Output per man can be written as $g(h)$; $h = H/L$. If g has no maximum, there will be excess demand for land at all share rental rates. If, on the other hand, g has a maximum at h_0, there will be excess demand for land from equation (16) for all share rents less than r_0:

$$(19) \qquad r_0 = 1 - w/g(h_0) = \operatorname*{Max}_h \left\{1 - w/g(h)\right\}$$

Either this will maximize the landlord's income per acre or the landlord is free to choose a lower value of r and ration the amount of land as in the first case. In both cases the landlord is free to set the share rent regardless of the decisions of other landlords. The conditions for this maximum are readily derived by maximizing rent per acre $rf(l)(= rg(h)/h)$ subject to equation (6) and take the neat form

$$(20) \qquad \frac{1 - r}{r} = \frac{1 - \rho}{\rho}\sigma$$

where σ is the elasticity of substitution between labor and land and ρ is the imputed share of land in the total product $(= HF_1/F)$.[1]

[1] $d/dr\ (rf(l)) = f + rf'\, dl/dr$; but $f' = w/(1-r)$, so $f''\, dl/dr = w/(1-r)^2 = f'/1-r$. Now $\sigma = -f'(f-lf')/lff''$, whence $(1-r)/r = f'^2/ff'' = lf'\sigma/(f-lf') = \sigma(1-\rho)/\rho$.

III. Existence and Stability of the Monopoly Equilibrium

The only equilibrium in which labor inputs by the tenant can be freely chosen has been shown to be a pure monopoly "equilibrium" in that each landlord can *entirely* ignore the decisions of all other agents and there is excess demand for land. Production is inefficient, and therefore below what it could be, while both tenants and landlords may be numerous and noncollusive. Any tenant who can persuade a landlord that he will reduce this inefficiency to the benefit of the landlord will clearly be preferred by the landlord—and the existence of inefficiency means that this is in principle possible at the same time making the tenant better off. The tenant's offer can take a variety of forms: a promise not to work for others, to work hard, to produce a satisfactory harvest, to accept dismissal if performance is not satisfactory, and so on. Such extensions of the contract will, if feasible and successful, replace the B-S equilibrium by an efficient equilibrium of the Cheung variety. Thus it can be argued that not only is the B-S equilibrium noncompetitive, but it may also be unstable, tending to be replaced by an efficient equilibrium.

IV. Parametric Shifts in the Wage Rate

Bardhan and Srinivasan use their model to predict an increase in the amount of land leased under share tenancy if the wage rate rises. Their rather complex algebraic argument is demonstrated in Figure 1, which claims to show the equilibrium share rent r^* and area of leased land supplied (q) and demanded (H). A rise in w is argued to raise the amount leased out at any given r from q_0 to q_1, say, and to lower the tenants' demand from $H_0(r)$ to $H_1(r)$. The net result is to lower the equilibrium rental share from r_0^* to r_1^*, and raise the amount leased from H_0^* to H_1^*. But this argument is invalid since we have seen that a competitive equilibrium of this nature does not exist. We can examine the predictions of the monopoly equilibrium analysis first under constant returns and then under diminishing returns.

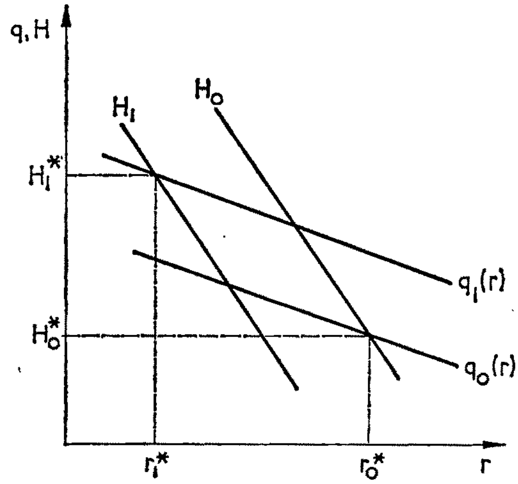


FIGURE 1. DEMAND AND SUPPLY IN THE MARKET FOR LAND LEASES

A. *Constant Returns*

The landlord will maximize his total income W by renting out H at the rental maximizing share rent $R=rf$ per acre.

$$W = G(A - H, L) - wL + HR(w) \tag{21}$$

so

$$G_1 = R(w) \tag{22}$$

$$G_2 = w \tag{23}$$

The effect of a parametric change in w is given by

$$\frac{dH}{dw} = -\begin{vmatrix} R_w & G_{12} \\ 1 & G_{22} \end{vmatrix} \div \begin{vmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{vmatrix} \tag{24}$$

The denominator is positive for a strictly concave function, while

$$\frac{dR}{dw} = \frac{\partial R}{\partial w} + \frac{\partial R}{\partial r}\frac{dr}{dw}$$

But $\partial R/\partial r=0$ as R is maximized with respect to r, (the familiar "envelope theorem" result) so, given $(1-r)f'(l)=w$,

$$\frac{\partial R}{\partial w} = \frac{r}{1-r}\frac{f'}{f''} < 0 \tag{25}$$

The sign of H_w is the sign of $(G_{12}-R_wG_{22})$

which is ambiguous, though quite likely to be negative in direct contrast to B-S. The reason is that the demand curves in Figure 1 are irrelevant, while an increase in w both lowers the attractiveness of the landlord farming his own land and the return he gets from renting it out. The final outcome will depend on the relative strength of the two effects and cannot be predicted a priori.

B. *Cropsharing with Diminishing Returns*

Suppose, however, that the tenant faces diminishing returns either because he needs to exercise scarce supervisory skill or because it takes him progressively longer to walk to far lying fields; in short because there is some essential indivisibility. This has the immediate consequence that a landlord will be concerned about the quantity of land the tenant hires from other landlords, since the intensity of cultivation (L/H) will decrease as the area cultivated increases.[2] We would not be surprised to find sharecroppers restricted to hiring land from one landlord, and such restrictions are quite common. Now consider the tenant renting in land from the landlord up to a maximum permitted amount $\overline{H}$, with equilibrium conditions given by equations (15) and (16) above. Suppose, moreover, that the demand curve is finite valued as shown in Figure 2, curve D. Anywhere to the left of D the tenant is keen to rent in more land no matter what the rental share, or, equivalently, the landlord is free to offer any value of r and H to the left of D. The only provision is that he has to offer a sufficiently attractive contract for the tenant to agree not to hire land from other landlords. Presumably this implies guaranteeing a minimum level of utility which, if the tenant is choosing to work elsewhere at the margin, implies both equation (15) and

$$(26) \qquad (1 - r)F(H, L) - wL \geqq C_0$$

Equation (26) implies a supply curve of tenants S which becomes horizontal at the

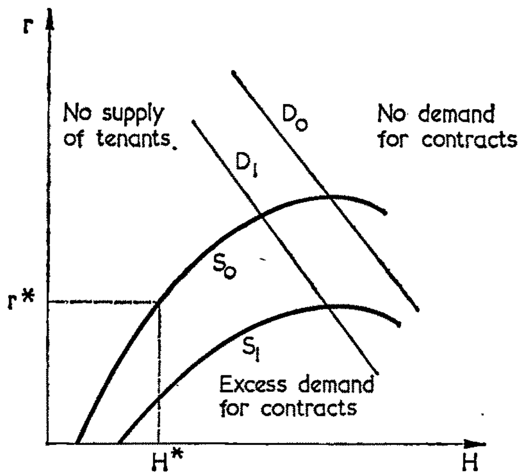


FIGURE 2. DEMAND AND SUPPLY UNDER CONTRACTUAL LIMITATION

point of intersection with D, as shown in Figure 2.[3]

The landlord's problem is to maximize his yield per acre rF/H subject to (15) and (26), at least if there is an elastic supply of identical tenants each facing diminishing returns, and he will offer tenancies lying on S, since rF/H decreases with H.[4] Unless the profit maximizing rental share is at the intersection of S and D, the landlord will maximize his income as a monopolist facing excess demand much as in the constant returns case, as for example in Figure 2 at the point r^*, H^*. If the wage rate w rises, or the minimum acceptable level of utility increases, the supply curve will shift to the right from S_0 to S_1, say, while a rise in w will also shift D to the left.

It is not obvious what the effect of a change in w will be on the total amount of land rented. Prima facie one would be surprised if the profit maximizing rent r^* increased sufficiently with a rise in w to offset the tendency for the supply of land available for leasing (q in Figure 1) to increase, but the link is not via shifts in the curve D,

[2] Differentiate (15) to obtain $d/dH(L/H) = -(HF_{12} + LF_{22})/H^2F_{22}$; but $HF_1 + LF_2 < F$ by Euler's theorem and diminishing returns, so $HF_{12} + LF_{22} < 0$ and the right side is negative.

[3] The equation for S is derived from (26) and (15) imposing equality. Its slope $\partial r/\partial H = (1-r)F_1/F \geqq 0$.

[4] $d/dH(rF/H) = r[HF_1 + HF_2\, dL/dH - F]/H^2$. But $dL/dH = -F_{12}/F_{22}$ from (15) so [ ] $= HF_1 - HF_2F_{12}/F_{22} - F < HF_1 + LF_2 - F < 0$.

but through shifts in S leading to a different r^* and hence a different number of tenants each renting a different area of land.

REFERENCES

P. K. Bardhan and T. N. Srinivasan, "Cropsharing Tenancy in Agriculture: A Theoretical and Empirical Analysis," *Amer. Econ. Rev.*, Mar. 1971, *61*, 48–64.

S. N. S. Cheung, "Private Property Rights and Sharecropping," *J. Polit. Econ.*, Nov./Dec. 1968, *76*, 1107–22.

———, *The Theory of Share Tenancy*, Chicago 1969.

D. M. G. Newbery, (1973a) "The Choice of Rental Contract in Peasant Agriculture," in L. Reynolds, ed., *Agriculture in Development Theory*, forthcoming.

———, (1973b) "Cropsharing Tenancy in an Equilibrium Model," paper presented to European Meeting of Econometric Society, Oslo 1973.

C. H. H. Rao, "Uncertainty, Entrepreneurship, and Sharecropping in India," *J. Polit. Econ.*, May/June 1971, *79*, 578–95.

L. Reynolds, *Agriculture in Development Theory*, New Haven, forthcoming.

Cropsharing Tenancy in Agriculture: Rejoinder

By P. K. Bardhan and T. N. Srinivasan*

The main argument of D. M. G. Newbery against the Bardhan-Srinivasan (B-S) model is that " . . . the B-S equilibrium requires a scarce resource (land) to be priced effectively at zero with no excess demand—a requirement which in general cannot be met" (p. 1063). Here Newbery confuses the price of land (service) with its marginal product. We contend that the relevant price of land is the return (in terms of consumption) that will accrue to the economy if the aggregate supply of land is increased by a small amount. In our model, the aggregate land supply was owned by landlords and was kept fixed at one unit. To determine the price, let us assume that the aggregate holding is α units. The aggregate consumption C is the sum of tenant's (C^1) and landlord's (C^2) consumption. For simplicity let us ignore with Newbery the income-leisure choice by tenants as well as landlords. Then, in B-S notation

$$C^1 = (1-r)F(H, l_1) - l_1 w$$

$$C^2 = G(\alpha - q, x) - wx + rF(q, L)$$

(Incidentally, Newbery is wrong in asserting that we *assume* G to be different from (and inferior to) F. We do not rule out G being identical to F in functional form.)

The price of land is $d(C^1+C^2)/d\alpha \equiv dC/d\alpha$ evaluated at the equilibrium values of H, r, l_1, q, x, and L. Now

$$\frac{dC}{d\alpha} = \frac{\partial C}{\partial H}\frac{dH}{d\alpha} + \frac{\partial C}{\partial r}\frac{dr}{d\alpha} + \frac{\partial C}{\partial l_1}\frac{dl_1}{d\alpha} + \frac{\partial C}{\partial q}\frac{dq}{d\alpha} + \frac{\partial C}{\partial x}\frac{dx}{d\alpha} + \frac{\partial C}{\partial L}\frac{dL}{d\alpha}$$

Now first-order conditions relating to maximization of C^1 and C^2 by the tenant and the landlord, respectively, ensures $\partial C/\partial H = \partial C/\partial l_1 = \partial C/\partial q = \partial C/\partial x = 0$. Hence $dC/d\alpha = \partial C/\partial r \cdot dr/d\alpha + \partial C/\partial L \cdot dL/d\alpha$. It is easily seen that $\partial C/\partial r \equiv 0$, $\partial C/\partial L = rF_2$, and $dL/d\alpha = dl_1/dr \cdot dr/d\alpha$ since in equilibrium $L = l_1(r)$. It can be shown that

$$\frac{dr}{d\alpha} = \frac{\partial q/\partial\alpha}{\dfrac{\partial H}{\partial r} - \dfrac{\partial q}{\partial r} - \dfrac{\partial q}{\partial L}\dfrac{dl_1}{dr}} < 0$$

since

$$\frac{\partial q}{\partial \alpha} = \frac{G_{11}G_{22} - G_{12}G_{21}}{rF_{11}G_{22} + G_{11}G_{22} - G_{12}G_{21}} > 0$$

and as shown (by B-S): $\partial q/\partial r = 0$, $\partial q/\partial L > 0$, $dl_1/dr < 0$ in equilibrium. Thus the equilibrium share r *decreases* as aggregate land availability increases. This ensures that $dC/d\alpha$, the price of land, is positive.

I

Newbery's statement that no competitive equilibrium can exist in the B-S model, derived on the basis of his equations (14)–(18), is inconsistent and wrong. In deriving (18) he uses our assumption that F is strictly concave, but then, contrary to his statement, (16) and (17) are *not* equivalent. It is easy to see that (18) may be valid and yet the interior solution in (16) may exist, as assumed by B-S. Even under constant returns to scale a competitive equilibrium can be shown to exist although we do not propose to do so here.

Since Newbery is wrong about the non-existence of competitive equilibrium in the B-S model, Sections III and IV of his paper are irrelevant.

Newbery is correct in pointing out that the brief discussion in our paper about parcellization implicitly assumes constant returns to scale, which is not in keeping with our explicit assumption of diminishing returns to scale in the rest of the paper.

* Professors of economics, Indian Statistical Institute, New Delhi.

II

In Newbery's paper where he formalizes S. N. S. Cheung's competitive theory of share tenancy, he finds that the constraint on tenant's use of labor intensity is binding (i.e., $l=l^*$) in equation (2). This, however, follows from his equation (3) only if one rules out zero marginal product of land. If zero marginal product of land (at a positive amount of l) is admissible, then under constant returns to scale (assumed by Newbery), it is possible to have $f=f'l$, in which case equation (3) implies $(1-r)f'=w$ in equation (2), which suggests inefficiency of sharecropping.

Newberry regards the model in Section I as *competitive*, yet in his equation (4) the landlord maximizes his income with respect to r, the crop share. It is a curious competitive model where each atomistic landlord takes the crop share as a variable of his choice. In case we do not allow the landlord this choice, so that he maximizes (4) with respect only to l^*, we do not get the efficiency condition (6) except in a special case. This can be shown as follows. Maximizing his equation (4) with respect to l^*, we get

$$f'(l^*)\left[\frac{r}{\lambda}+(1-r)\right]=w$$

$$=\frac{(1-r)f(l^*)}{l^*}$$

from Newbery's (3). Now from this to derive Newbery's efficiency condition (6), one needs $r/\lambda+(1-r)=1$ implying either $r=0$ or $\lambda=1$. As long as $r\neq 0$ in the sharecropping equilibrium, this implies $\lambda=1$, or

$$\frac{(1-r)f(l^*)}{f'(l^*)l^*}=1$$

or

$$r=\frac{f(l^*)-f'(l^*)l^*}{f(l^*)}$$

= land elasticity in the production function

Except in this very special case Newbery will not get his efficiency condition (6).

III

In our paper we assumed that the landlord cannot decide how much labor the sharecropper puts in his land. Newbery formalizes Cheung's model by assuming instead that the landlord specifies a minimum labor intensity on the sharecropped land. In justification Newbery cites Cheung's examples (given on pp. 76–77 of his book) of actual share contracts in pre-Communist China where the amount of tenant labor (in terms of "so many bodies of men") as well as the rental share is specified in the contract. However, Cheung's examples are by no means conclusive. After all, in any legal contract the obligations of each of the contracting parties are specified. But to infer from this that one of the parties (the landlord in this case) was the party who determined the terms of the contract may be unwarranted. A contract is the *ex post* outcome of a negotiation, it cannot prove the lack of *ex ante* choice on the part of the tenant. Then again, specification of the number of "bodies of men" does not ensure the actual quantity of labor hours, not to speak of the quality.

IV

In conclusion we would like to acknowledge that the B-S model is not fully satisfactory, not only for the reasons mentioned in the original article. For instance, the definition of competitive equilibrium we adopted was based on treating the crop share r as a price-like variable, and it was postulated that tenants and landlords treat it as a parameter. But it is not a price in the sense that it represents the exchange value of a unit of land in terms of some numeraire. The fact that behavior in which prices are treated as parameters by all participants may at times lead to inefficiencies is not, however, unfamiliar (as for instance in the consumption loan model of Paul Samuelson or the storage example of Tjalling Koopmans).

REFERENCES

P. K. Bardhan and T. N. Srinivasan, "Crop-sharing Tenancy in Agriculture: A Theoretical and Empirical Analysis," *Amer. Econ. Rev.*, Mar. 1971, *61*, 48–64.

S. N. S. Cheung, *The Theory of Share Tenancy,* Chicago 1969.

T. C. Koopmans and A. F. Bausch, "Selected Topics in Economics Involving Mathematical Reasoning," *SIAM Rev.,* July 1959, *1,* 79–148.

D. M. G. Newbery, "Cropsharing Tenancy in Agriculture: A Theoretical and Empirical Analysis: Comment," *Amer. Econ. Rev.,* Dec. 1974, *64,* 1060–66.

P. A. Samuelson, "An Exact Consumption Model of Interest with or without the Social Contrivance of Money," reprinted in J. E. Stiglitz, ed., *The Collected Scientific Papers of Paul A. Samuelson,* Vol. I, Cambridge 1966, 219–34.

Incentive Contracts and Competitive Bidding: Comment

By COLIN C. BLAYDON AND PAUL W. MARSHALL*

In a recent article in this *Review*, David Baron concludes that optimal bids for incentive contracts would decrease with increased risk aversion. This is a dramatically counterintuitive result and deserves further comment.

The profit on an incentive contract, if won, is directly proportional to the bid. So a natural first thought is that the more risk averse the bidder, the higher the bid. However, Baron examines a case where the cost of completing the contract, if won, is known with certainty. Thus, the only uncertainty in the problem is uncertainty about what the competition will bid. Since the low bid wins the contract, it is plausible in this case for the more risk-averse bidder to bid lower in order to get a higher probability of winning, and in fact, this is the result that Baron proves. He then says that the same result holds in the case where the cost to complete the contract is uncertain. This is not so plausible and can be shown to be not true in some cases.

Uncertain costs make for uncertain profits and this should motivate the more risk-averse contractor to bid higher. But, as we just noted, this motivation is countered by the risk of losing the contract, which forces a contractor to lower his bid to get a higher probability of winning. All this indicates that any conditions which specify the impact of increased risk aversion on bidding behavior will have to involve tradeoffs between the probability of winning the contract, the uncertainty of the costs, and the degree of risk aversion. A simple example illustrates this. Suppose the cost to complete the contract is normally distributed with mean $\bar{c}$ and variance σ^2. The contractor exhibits constant risk aversion with a utility for profits of the form

$$-e^{-\lambda(\text{profits})}$$

The contractor assesses the lowest competitive bid $\tilde{p}$ to be distributed exponentially with a density function ke^{-kp}. For these assumptions, the optimal bid $\hat{p}$ is given as[1]

$$(1)\quad \hat{p} = \left(\frac{1}{\alpha+\beta}\right)\left[\frac{1}{\lambda}\, ln\left(1+\frac{\lambda(\alpha+\beta)}{k}\right) + \beta\bar{c} + \frac{\lambda\sigma^2\beta^2}{2}\right]$$

How does this bid vary as the degree of risk aversion (as measured by the parameter λ) increases? For the case where costs are certain (i.e., for $\sigma^2=0$), $\hat{p}$ decreases as λ increases just as we would expect. But when costs are uncertain it is not clear whether $\hat{p}$ increases or decreases. If we differentiate $\hat{p}$ with respect to λ, we get

$$(2)\quad \frac{d\hat{p}}{d\lambda} = \left(\frac{1}{\alpha+\beta}\right)\left[\frac{\sigma^2\beta^2}{2} + \frac{1}{\lambda^2}\cdot\left\{\frac{\lambda(\alpha+\beta)/k}{1+\frac{\lambda(\alpha+\beta)}{k}} - ln\left(1+\frac{\lambda(\alpha+\beta)}{k}\right)\right\}\right]$$

The condition for $d\hat{p}/d\lambda$ to be positive, i.e., for the optimal bid to increase with increased risk aversion is:

$$(3)\quad \frac{\sigma^2\beta^2}{2} > \frac{1}{\lambda^2}\left[ln\left(1+\frac{\lambda(\alpha+\beta)}{k}\right) - \frac{\lambda(\alpha+\beta)/k}{1+\frac{\lambda(\alpha+\beta)}{k}}\right]$$

* Deputy associate director, Office of Management and Budget, on leave from the Graduate School of Business Administration, Harvard University, and associate professor, Graduate School of Business Administration, Harvard University, respectively.

[1] In Baron's notation, α is the percentage fee and β is the sharing ratio for cost overruns and underruns so profit is $\alpha p+\beta(p-c)$.

This condition involves parameters that characterize the contractor's risk aversion (λ), the uncertainty in costs (σ^2), and the competitiveness of other bidders (k). It can be seen immediately that the condition can be met if the variance of the cost, σ^2, is large enough. For such cases, the optimal bid will increase with increased risk aversion.

The optimal bid in this example behaves just as one would expect in other ways as well. As costs get more uncertain (σ^2 increasing), the optimal bid goes up and as other bidders get more competitive (k increasing), the optimal bid goes down.

This example illustrates, then, that Baron's Proposition 1 does not hold, in general, for cases where the cost of performing the contract is uncertain. In particular, for uncertain costs, the condition analogous to Baron's condition (8) for certain costs is that[2]

$$(4)\quad -f(\hat{p}_2)\left[\frac{E_c\{u_2(w+\pi(\hat{p}_2)+R)\}-u_2(w+R)}{E_c\{u_2'(w+\pi(\hat{p}_2)+R)\}} - \frac{E_c\{u_1(w+\pi(\hat{p}_2)+R)\}-u_1(w+R)}{E_c\{u_1'(w+\pi(\hat{p}_2)+R)\}}\right]\leq 0$$

The fact that u_1 is everywhere more risk averse than u_2 is not sufficient by itself for the above condition to be satisfied as it is in the case where costs are certain.

REFERENCE

D. P. Baron, "Incentive Contracts and Competitive Bidding," *Amer. Econ. Rev.*, June 1972, *62*, 384–94.

[2] In Baron's article, w represents the firm's wealth, R is profit from private sector operations, $\pi(\hat{p}_2)$ is the profit from the contract when the bid is $\hat{p}_2$ and $\hat{p}_2$ is the optimal bid for a contractor with the utility function u_2.

Incentive Contracts and Competitive Bidding: Reply

By David P. Baron*

In their comment, Colin Blaydon and Paul Marshall have indicated that while greater absolute risk aversion results in a lower optimal bid price for incentive contracts as stated in my Proposition 1,[1] the result may not hold for the case in which project costs are uncertain. The purpose of this reply is to interpret and to extend the analysis of this point.

A firm may face two risks associated with the opportunity to bid on a contract, and the effect of differences in risk aversion on the optimal bid price is composed of parts corresponding to each risk. The first effect pertains to the risk that the firm will or will not be awarded the contract, and for a given project cost this will be referred to as the bid effect. The second pertains to the risk associated with an uncertain project cost, conditional on being awarded the contract, and this will be referred to as the cost uncertainty effect. To examine these two effects, the necessary optimality condition with a utility function U_2 will be written as

$$(1)\quad (1-F(\hat{p}_2))U_2'(x_2)(\alpha+\beta)(1-\Delta\Pi_{2C}') - f(\hat{p}_2)(U_2(x_2)-U(v)) = 0$$

where $x_2=w+\alpha\hat{p}_2+\beta(\hat{p}_2-\overline{C})-\Delta\Pi_{2C}+R, v=w+R$; w is wealth, α is the target profit rate, β is the incentive sharing rate, R is profit from other activities, $\overline{C}$ is the expected project cost, $\hat{p}_2$ is the optimal price, $\Delta\Pi_{2C}$ is the risk premium corresponding to the uncertain cost, $\Delta\Pi_{2C}'$ is the marginal risk premium, and F and f are the distribution and density functions, respectively, for the low bid by competitors. Proposition 1 in my article indicated that the bid effect reduces the optimal bid price when project costs are certain. The essential step in the proof of that result utilizes the property that if a utility function U_1 is more absolute risk averse than U_2 (see John Pratt, equation (22)), then

$$(2)\quad \frac{U_2(x)-U_2(v)}{U_2'(x)} < \frac{U_1(x)-U_1(v)}{U_1'(x)}$$

if $x>v$. Substituting from (2) evaluated at $\hat{p}_2$ and x_2 into the necessary optimality condition (1) for U_2 yields

$$(3)\quad (1-F(\hat{p}_2))U_1'(x_2)(\alpha+\beta)(1-\Delta\Pi_{2C}') - f(\hat{p}_2)(U_1(x_2)-U_1(w+R)) < 0$$

This effect for a deterministic cost (and hence $\Delta\Pi_{2C}=\Delta\Pi_{2C}'=0$) implies that the optimal price $\hat{p}_1$ for U_1 is less than $\hat{p}_2$ for U_2, given that expected utility is strictly concave.

With uncertain costs, however, the certainty equivalent of the profit on the contract (conditional on the contract being obtained), is decreasing in risk aversion where the certainty equivalent is defined by $CE_i(p)=\alpha p+\beta(p-\overline{C})-\Delta\Pi_{iC}$. Since the left side of (3) is decreasing in the certainty equivalent and $CE_1(\hat{p}_2)<CE_2(\hat{p}_2)$, depending on the relationship between $\Delta\Pi_{2C}'$ and $\Delta\Pi_{1C}'$, the cost uncertainty effect can tend to increase the optimal bid price.[2] If the cost uncertainty effect offsets the bid effect, an increase in risk aversion will result in a higher optimal bid price.

Differences in risk aversion affect the optimal bid price, which in turn affects the optimal certainty equivalent. For the example considered by Blaydon and Marshall, the optimal conditional certainty equivalent is $CE=(1/\lambda)ln(1+\lambda(\alpha+\beta)/k)$, which is independent of the probability distribution of cost and is decreasing in the index of absolute

* Professor of managerial economics and decision sciences, Northwestern University.

[1] The statement and proof of Proposition 1 contain typesetting errors not detected in the galleys. The second sentence in the statement of the proposition should read "Then $\hat{p}_1\leq(<)\hat{p}_2$." Also, in line three of the second column of page 387, the signs should be $\geq(>)$.

[2] In the example of Blaydon and Marshall $\Delta\Pi_C'=0$.

risk aversion.[3] In this example the firm finds it optimal to adjust its price in response to differences in risk aversion in such a manner that the resulting conditional certainty equivalent of the contract is independent of the probability distribution of cost.[4] Since the optimal conditional certainty equivalent is positive, the firm finds it desirable to bid given any distribution of cost.

The effect of risk aversion differences on the optimal bid price also depends on the sharing rate β. One extreme is the cost-plus-fixed-fee contract ($\beta=0$) for which the conditional profit is independent of the project cost. In that case, greater absolute risk aversion results in a lower bid price for any distribution of project cost. The risk aversion effects also depend on the "degree of uncertainty" regarding project costs. For contracts involving considerable cost uncertainty, such as Department of Defense development-production contracts, contractors prefer low values for the sharing rate and typically contracts are let by a process involving both bidding and negotiation.

[3] $1/k$ is the mean of an exponential distribution $F(p)$ and λ is the index of absolute risk aversion for the constant absolute risk averse utility function $U(y)=-\exp(-\lambda y), \lambda>0$. The distribution of cost is assumed to be normal.

[4] The optimal bid price however is an increasing function of the variance of cost.

REFERENCES

D. P. Baron, "Incentive Contracts and Competitive Bidding," *Amer. Econ. Rev.*, June 1972, *62*, 384–94.

C. C. Blaydon and P. W. Marshall, "Incentive Contracts and Competitive Bidding: Comment," *Amer. Econ. Rev.*, Dec. 1974, *64*, 1070–71.

J. W. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica*, Jan./Apr. 1964, *32*, 122–36.

Feige and Parkin on the Optimal Quantity of Money

By THOMAS RUSSELL*

Edgar Feige and Michael Parkin have recently made a clear statement in this *Review* of the conditions under which the paying of interest on real money balances is socially optimal. The purpose of this note is to show that their analysis is deficient in two places and that the problem of the optimal quantity of money may be more complex than they suggest.

The first problem that arises concerns their treatment of the individual budget constraint. An individual is to maximize a utility function $U(pq)$ defined over the consumption of a standard commodity subject to a budget constraint on his income account (see Feige and Parkin's equation (2)),

$$0 = Y + \pi - pq - T \tag{1}$$

where Y = labor income in current dollars
π = net profit from inventory management
pq = expenditure on commodities
T = taxes

The individual is also subject to a budget constraint on capital account (see Feige and Parkin's equation (8)),

$$\bar{A}^* = P\bar{K} + \bar{B} + \bar{M} + P\bar{Q} \tag{2}$$

where $\bar{A}^*$ = average fixed stock of assets
$P\bar{K}$ = average stock of capital in current dollars
$\bar{B}$ = average inventory of bonds in current dollars
$\bar{M}$ = average inventory of cash balances
$P\bar{Q}$ = average inventory of commodities in current dollars

Equations (1) and (2) differ from the standard Tobin-Baumol model in two ways. First, the range of inventories is larger, being extended to include goods and capital; and second, the income earned on portfolio management is treated as spendable across the planning period. Neither of these additions necessitates any fundamental change in the analysis, however, and although the system of necessary conditions would have to be solved recursively, one could in principle derive a demand function for bonds, money, goods, and capital as functions of the exogenous prices and transaction costs in the normal way.

This, however, is not the way Feige and Parkin proceed. One obvious problem with treating capital as another income earning asset in a one-period framework is that the individual would never plan to have any capital at the end of the period. Feige and Parkin attempt to overcome this problem as follows. First of all they exclude the possibility that individuals may hold capital as a temporary abode of purchasing power. Secondly they introduce the equation (see their equation (19)),

$$\bar{W}^* = P\bar{K} + P\bar{Q} \tag{3}$$

which they justify as follows: "The final constraint derived from the stationary assumption, states that society's stock of inventories of physical capital plus the stock of commodity inventories is constant" (p. 342).

Now this equation is, in effect, Feige and Parkin's demand for capital equation, but since it is defined implicitly it is not easy to see it as such. In effect what the individual does is as follows. He tries out some level of expenditure pq. If he arranges his holdings of money, bonds, and goods in an optimal fashion, this determines $P\bar{Q}$. It therefore also determines by (3), $P\bar{K}$. This however is not the end of the story since it also determines π, and therefore at the given pq, equation (1) may not be satisfied. Thus pq must be adjusted to meet the budget constraint. This is a recursive process which converges at pq^* with $P\bar{K}^*$ being given by $\bar{W}^* - P\bar{Q}^*$. Thus capital in this model in no way competes

* University of Essex.

with money or bonds as a carrier of purchasing power. Rather capital is a residual whose only property in equilibrium is that the interest earned on it balances the consumer budget. This, however, can lead to most perverse results. Consider first $\partial P\overline{K}/\partial r_K$ where r_K is the return on capital. It is easy to show (see the Appendix) that whereas $\partial P\overline{K}/\partial r_K$ is in general unsigned, at the particular values of Feige and Parkin's numerical example it is actually negative. That is, paying a higher rate of return on capital causes less capital to be held. It is also shown by Feige and Parkin themselves that the effect of raising r_m is to cause more physical capital to be held. Neither of these properties of the demand for capital function seems to make economic sense. Capital, just like bonds or money, is a provider of nonpecuniary returns. It must therefore compete with bonds and money for a position in the individual's portfolio. To say $\partial P\overline{K}/\partial r_K$ is less than zero and $\partial P\overline{K}/\partial r_m$ is greater than zero is to deny any such function to capital. Indeed it is also to deny any function to financial intermediation, for the holding of nonliquid assets (capital) in the Feige and Parkin model actually increases as the spread between the return on liquid assets (money) and nonliquid assets shrinks. Surely it is more reasonable to argue that as the rate of interest on money r_m approaches that on capital r_K, the demand for capital falls as individuals substitute the liquid asset for it. If this were true, however, Feige and Parkin's argument would fall to the ground, for then the paying of interest on money would cause individuals to hold less of the private sector's only productive asset. To the extent that money is outside money, savings would now be invested with the state. To the extent that the money is inside money, there would be a blockage in the flow of funds, and banks would simply deposit with each other. In either case it is clear that if Feige and Parkin wish private real capital to be held by individuals, some spread must be offered between the rates of return on capital and money in order to compensate for the loss of liquidity. Just what his spread must be will depend on individual preferences and the efficiency of financial intermediation. It is perfectly true to say that whatever the individual does not hold as a stock of commodities is *available* for investment. It is quite another matter however to show how these savings get channelled into investment, and the r_m, r_K structure which guarantees any given volume of investment is clearly quite a difficult matter to calculate. In any case we should reject completely any portfolio model which suggests that as we reduce the spread between the rates on nonliquid and liquid assets, more nonliquid assets are held.

The second problem with Feige and Parkin's analysis is their treatment of the stock of commodity inventories. Now it is quite clear that the paying of an interest rate on money reduces the *individual* consumer's demand for inventory of commodities. What Feige and Parkin are required to show, however, is that the paying of interest on money balances reduces *society's* stock of commodity inventories. It is surely possible that the reduction in consumer inventories is just offset by an increase in shop or wholesale inventories. Now to answer the question of what happens to society's stock of commodity inventories when we pay interest on money, we must know more about the nature of the production process. Suppose for example we had conveyor belt production with output coming off the production line continuously during the period of discussion. Then clearly the reduction of consumers' commodity inventories reduces society's inventories. However, we already know that production in Feige and Parkin's world cannot be continuous. If it were, it would not be possible in the case where no interest is paid on money for consumers to buy pq/m commodities on the first trip to the shops where m is the number of trips to the shops. Clearly this production must already have taken place. Thus if individuals no longer wish to hold it as inventories, someone else in society must hold it. Thus on this first batch of production there is no social saving whatsoever in terms of reduced inventories of commodities. In order to calculate precisely the *social* change in inventories, we must know

exactly the pattern of production. From Feige and Parkin's point of view, the worst that could happen would be that all output was produced on the first day of the period. Then the social saving would be precisely zero. This represents an extreme position but once again we see that to calculate the social gains from paying interest on money we need more information than that provided by Feige and Parkin.

The great virtue of Feige and Parkin's treatment of the problem of monetary optimality is that it states clearly the model of discussion. However, just because it is so clear, the gaps in the analysis are the more clearly revealed. It has been shown that a) the demand for capital in the Feige-Parkin model has very unreasonable properties, and b) the treatment of commodity inventories is a partial treatment and does not consider the problem from the point of view of society. If the model is extended to include a realistic demand for capital function and a full treatment of inventories, the "correct" rate of interest on money would seem to become a much more complex problem.

Appendix

The Sign of $\partial(P\overline{K})/\partial r_K$

From equation (3) (Feige-Parkin's equation (19)),

$$P\overline{K} = \overline{W}^* - \frac{pq}{2m}$$

where $m = \sqrt{(r_m - \alpha_m + \alpha_q)/2_{\beta q}(pq)}$ and r_m, α_m, β_q, and q are as defined by Feige and Parkin. Thus clearly $\partial(P\overline{K})/\partial r_K$ has the opposite sign from $\partial(pq)/\partial r_K$. Now pq is given by Feige and Parkin's equation (20)

$$(pq)^* = \frac{X \pm \sqrt{X^2 - 4A^2B^2}}{2B^2}$$

Thus

$$\frac{\partial(pq)^*}{\partial r_K} \quad \left\{\frac{\partial X}{\partial r_K}(X^2 - 4A^2B^2)^{1/2} \pm \left[X\frac{\partial X}{\partial r_K} - 4B^2A\frac{\partial A}{\partial r_K}\right]\right\} \div [2B^2(X^2 - 4A^2B^2)^{1/2}]$$

This expression is in general unsigned but it equals 4962 approximately where X, A, and B are evaluated at the values of Feige and Parkin's numerical example. Thus in this case $\partial(P\overline{K})/\partial r_K$ is negative.

REFERENCE

E. L. **Feige and M. Parkin,** "The Optimal Quantity of Money, Bonds, Commodity Inventories, and Capital," *Amer. Econ. Rev.*, June 1971, *61*, 335–49.

Population Growth, Agricultural Capital, and the Development of a Dual Economy

By YOSHIO NIHO*

David Ricardo and Thomas Malthus saw the problem of population growth and attendant diminishing returns to labor as a major economic problem. Diminishing returns to labor led, of course, to the gloomy Malthusian law of population growth limited only by the subsistence wage of labor. In contrast, neoclassical growth theory which is characterized by its optimistic outlook on an economy's ability to achieve balanced growth shunts aside the problem of population growth and diminishing returns to labor as a mere qualification to the main argument. The difference between the conclusions of classical and neoclassical growth theories stems from their assumptions concerning the production function. In classical theory, land enters the production function as a factor limiting the level of output; its fixity in supply then characterizes the production function by diminishing returns to scale. The neoclassical conclusion of a stable balanced growth, on the other hand, is based on the assumption that production processes proceed by constant returns to scale with respect to capital and labor; land does not enter the production function as a factor limiting the level of output.

The neoclassical conclusion may be valid in a situation in which land is a free good: its supply increases in proportion to other factors at zero price. The United States in the early nineteenth century represented such a situation when the frontier was defined simply as uncleared land. The neoclassical conclusion may also be applied to a highly industrialized state of an economy in which the proportion of agriculture is small and in which the agricultural sector is highly modernized. In today's less developed countries, however, a traditional agricultural sector is dominant in the national economy. For such an economy, classical theory, rather than neoclassical, is the more relevant.

Conventional theories of development, such as those developed by Arthur Lewis, John Fei and Gustav Ranis, Dale Jorgenson, Paul Zarembka, Ryuzo Sato and the author, have specified the agricultural production condition by the classical assumption of diminishing returns to scale.[1] Conclusions derived by these studies clarified the problem of development as one of the balance between two forces: productivity increases in agriculture caused by various forms of technical progress; and the forces of diminishing

* Assistant professor of economics, University of Wisconsin-Milwaukee. This paper is taken from my doctoral dissertation. I am especially indebted to my adviser Ryuzo Sato for his numerous helpful comments and constant encouragement. Also, I would like to thank George Borts, James Hanson, Tong Hun Lee, John Makin, Melvin Lurie, Ulaganathan Sankar, Jerome Stein, and an anonymous referee for valuable criticism and encouragement.

[1] Sometimes conventional development theories are classified into two categories: "classical" and "neoclassical." The two branches of theory differ in the assumption regarding whether disguised unemployment is existent or not, and in the assumption regarding whether the real wage rate in agriculture is constant or variable. These differences become important when one tries to identify which theory is more applicable to a particular economy. This must be determined by empirically testing the two models. However, when one discusses the conditions of steady development, these differences become immaterial. Under either set of assumptions, the condition for the sustained growth of per capita income and of the industrial sector is shown to be a rate of technical progress in agriculture that is sufficient to offset the force of diminishing returns to labor resulting from population growth. Thus, in this paper no distinction is made between the two theories. I refer to both branches of existing theory as "conventional," in the sense that it is common in both theories to exclude capital from the agricultural production function and consequently to have productivity conditions in industry independent from the conditions of development. (One exception is the work of Zarembka, who discussed the case in which the price elasticity of the demand for food is not zero.)

returns to labor caused by population growth. The conditions necessary for development are summarized as the existence of a positive and growing agricultural surplus (see Jorgenson) or a maximum sustainable rate of population growth that exceeds the actual rate of population growth (see Sato and the author).[2]

In these conventional theories, however, capital was not included in the agricultural production function as an input;[3] agricultural output was assumed to be determined by the input of labor alone. This may have been due to the fact that these theories were intended to represent the traditional agricultural sector of contemporary developing countries. On the other hand, considerable historical evidence suggests that the application of capital inputs (such as fertilizer, insecticides, and machinery) in the agricultural sector was an important element in the successful development of modern economies such as Germany, Japan, and the United States.[4]

Indeed, the conclusions of conventional theories suggest that capital investment in agriculture is likely to be important since it may help to raise the rate of increase of the agricultural surplus or, equivalently, the maximum sustainable rate of population growth. Traditional arguments for the uses of labor-intensive techniques in a relatively labor-abundant situation do not seem valid for development. Such factor substitution is a stabilizing factor in a neoclassical economy, but in the context of development it becomes a stagnating factor. On the other hand, the use of more capital-intensive techniques in agriculture will raise the productivity of labor, thereby enabling the agricultural sector to release more labor to industry and to supply a larger amount of the wages fund. In addition, the introduction of capital into agriculture is likely to infuse modern knowledge and modern spirit into what is frequently a tradition oriented sector, thereby stimulating improvement in the quality of labor.

In this paper I introduce agricultural capital into the conventional model of development (see Sato and the author) in order to examine the effect of agricultural capital upon economic development. It is shown that the introduction of capital into agricultural production significantly raises the society's maximum sustainable rate of population growth, thus creating a much greater possibility of steady development for the economy. In this model the maximum sustainable rate of population growth is determined not only by the rate of technical progress in agriculture but also by the rate of (Harrod-neutral) technical progress in industry. Thus, the introduction of capital into agriculture adds another policy dimension to the model, giving us three alternatives available to achieve steady development under given production conditions; namely, attempts to control population growth, policies implemented to increase the rate of technical progress in agriculture, and measures designed to accelerate the rate of increase in the efficiency of labor in industry. This is a significant difference from the implications of conventional theories of development in which the technological and production conditions in industry cannot play any role in assisting the economy to escape the trap of stagnation.[5]

In the general model to be presented here, the conventional model of development and

[2] Actually, the two conditions imply the same thing. The difference between the maximum sustainable rate and the actual rate of population growth represents the rate of growth of agricultural surplus.

[3] In their model of a dual economy, Allen Kelley, Jeffrey Williamson, and Russell Cheetham introduced capital into agricultural production, but excluded land.

[4] Capital intensity in agriculture in the three countries was: Germany, 2.95 in 1850 and 5.41 in 1913 (both per man and in thousand marks in 1913 prices); Japan, 155 in 1874 and 232 in 1936 (both per man and in 1934–36 yen); United States, .644 in 1889 and .926 in 1936 (both per man-hour and in 1929 dollars.) See the author (1973b) for details.

[5] This implication of conventional theories is based on the assumption that the price elasticity of demand for food is zero. If it is not zero, the productivity conditions in industry will affect the maximum sustainable rate of population growth even if capital is not introduced into agricultural production. The case in which the price elasticity of food is not zero is studied by Zarembka with a model in which the population growth is exogenous and capital is not included in agricultural production.

neoclassical two-sector growth model are included as special cases. If the production elasticity of capital in agriculture is zero, this model is reduced to the conventional model of development (see Sato and the author). If the production elasticity of land is zero so that the agricultural production function is subject to constant returns to scale, the model reproduces the results of the neoclassical two-sector growth model. It is shown that if land is not a limitational factor in agricultural production, the growth of the population never works as a pressure upon the growth of per capita income; and that steady development is always achieved regardless of the behavior of the population growth, provided that there is positive technical progress in at least one sector. Thus, the result of the neoclassical two-sector growth model in which balanced growth is always generated with Cobb-Douglas type of production functions arises because land does not enter the production function as a factor restricting the level of output.

I. The Model

The production function in agriculture is assumed to be:

$$Y_a = A(t)K_a^{\beta}L_a^{\delta}N^{1-\beta-\delta}$$

where Y_a represents agricultural output, K_a capital stock in agriculture, L_a labor input in agriculture, N the amount of arable land, and $A(t)$ the technological improvement factor in agriculture. The coefficients β, δ, and $1-\beta-\delta$ are the production elasticities of capital, labor, and land, respectively, and they are assumed to be positive constants less than unity. Thus, the function exhibits constant returns to scale with respect to capital, labor, and land. If we assume that the amount of land is fixed and that technological improvement takes place at a constant annual rate, we can rewrite the production function in the form:

$$(1) \qquad Y_a = A_0e^{\alpha t}K_a^{\beta}L_a^{\delta}, \qquad \beta + \delta < 1$$

where A_0 represents the initial value of the technological improvement factor and the fixed amount of land, and α represents the constant annual rate of technical progress. Without loss of generality we can assume that A_0 is equal to unity. Due to the fixity of land, the agricultural production function is characterized by diminishing returns to scale.

The production function in industry is assumed to be:

$$(2) \qquad Y_m = B_0e^{\lambda t}K_m^{\gamma}L_m^{1-\gamma}$$

where Y_m represents industrial output, K_m capital stock in industry, L_m labor input in industry, and λ the constant annual rate of technical progress in industry; B_0 is the initial value of the technological improvement factor in industry and is assumed equal to unity; γ is the production elasticity with respect to capital and is assumed $0<\gamma<1$. Thus, the industrial production function exhibits constant returns to scale with respect to capital and labor.

The introduction of capital into agricultural production makes doubtful the prolonged existence of surplus labor in the agricultural sector. Infusion of capital into agricultural production will raise the marginal productivity of labor in the agricultural sector. If the real wage in the agricultural sector is determined by institutional forces, that is, independently of the marginal productivity of labor, continuous infusions of capital will eventually raise the marginal productivity of labor to the level of the institutionally determined real wage, with the concomitant depletion of surplus labor (see Fei and Ranis). Together with capital may come the spirit of capitalism, the pervasive dominance of profit maximizing considerations replacing institutional forces in the determination of real wages. Surplus labor is unlikely to persist in such an environment. However, in order to treat the case more generally, we assume that a constant fraction of agricultural output is paid as the total wage bill to the agricultural workers. Thus,

$$(3) \qquad \frac{w_a}{q} = \xi\frac{Y_a}{L_a}$$

where w_a is agricultural wage measured in terms of industrial goods, q the terms of trade (the price of agricultural goods in terms of industrial goods), and ξ the constant relative share of labor in agriculture which may or may not be equal to δ, the production elasticity of labor in agriculture.

The industrial real wage is assumed to be determined by the marginal productivity of labor in industrial production. Thus, if w_m denotes the industrial real wage, we have

$$w_m = (1 - \gamma)\frac{Y_m}{L_m} \tag{4}$$

where $1-\gamma$ is the relative share of labor in the industrial sector.

The marginal productivity of capital in each sector is assumed to determine its return to capital. Thus, if r_a is the return to capital in agriculture measured in terms of industrial goods and r_m is the return to capital in industry, we have

$$\frac{r_a}{q} = \beta\frac{Y_a}{K_a} \tag{5}$$

and

$$r_m = \gamma\frac{Y_m}{K_m} \tag{6}$$

where β and γ are the relative shares of capital in the agricultural and industrial sectors, respectively.

The total capital stock, K, is assumed to be distributed between the agricultural and industrial sectors such that its returns measured in terms of industrial goods in the two sectors are equalized. Thus, we have

$$K = K_a + K_m \tag{7}$$

and

$$r_a = r_m \tag{8}$$

The following relation is assumed to hold between the real money wages of the agricultural and industrial sectors:

$$w_a = \mu w_m \tag{9}$$

where μ is a constant and $0<\mu\leq 1$.

We assume that the total supply of labor, i.e., the labor force L, is a constant fraction of the total population P, and that L is divided into the agricultural and industrial labor forces. Thus, we have

$$L = \theta P \tag{10}$$

where θ is a constant and $0<\theta<1$, and

$$L = L_a + L_m \tag{11}$$

With regard to the supply of labor we assume the Malthusian population growth function, i.e.,

$$\frac{DP}{P} = \phi(z), \qquad \phi'(z) > 0 \tag{12}$$

where z is the level of per capita income measured in terms of agricultural units and is defined as

$$z = \frac{Y_a + Y_m/q}{P} \tag{13}$$

The demand function for agricultural output is assumed to be explained by Engel's Law.[6] Thus, if v denotes per capita food consumption, we have

$$v = \frac{Y_a}{P} \tag{14}$$

and

$$v = v(z), \quad 0 < v'(z) < 1, \quad v''(z) < 0 \tag{15}$$

In much of the literature on development theory, savings behavior is explained by the "classical" savings function, which states that wage earners do not save, while all profits are automatically saved. This assumption may be consistent with a situation in which wage incomes barely exceed the subsistence level. In this paper, however, we assume a savings function which is more common in a neoclassical growth theory, i.e., that the amount of total savings is a constant fraction of the gross national product. This is because in countries such as the United States, Germany, and Japan, where the application of agricultural capital was an

[6] The price elasticity of demand for food is assumed to be zero.

important magnitude, savings out of wage incomes may have been fairly sizable. As has already been noted, surplus labor is unlikely to persist in an economy in which capital is introduced into agricultural production. Thus, if S denotes the amount of total savings measured in terms of industrial goods, we have

$$S = \rho[Y_m + qY_a] \tag{16}$$

where ρ represents the constant savings ratio. The society's capital accumulation then is explained by

$$DK + \eta K = S \tag{17}$$

where ηK represents the amount of total depreciation.

II. Workings of the Model

Defining overall, agricultural, and industrial capital intensities $k=K/L$, $k_a=K_a/L_a$, $k_m=K_m/L_m$, and average productivities of labor in the agricultural and industrial sectors $y_a=Y_a/L_a$ and $y_m=Y_m/L_m$, the relationship between the labor productivities in the two sectors and the relationship between the capital intensities in the two sectors can be shown as:

$$y_a = C\frac{y_m}{q}$$

where $C=\mu(1-\gamma)/\xi=$ constant, and

$$k_a = Ek_m$$

where $E=\mu(1-\gamma)\beta/\gamma\xi=$ constant. It can be shown (see Appendix A) that the proportion of agricultural workers to the total labor force $s=L_a/L$ is determined as

$$s = s(z) = \frac{v(z)}{v(z) + C(z - v(z))} \tag{18}$$

Equation (18) implies[7] $0<s<1$ and

$$s'(z) = \frac{-C(v - v'z)}{[v + C(z - v)]^2} \leqq 0$$

[7] For the primitive agrarian economy which is engaged only in agricultural production, we have $z=v(z)$, $s=1$, and $s'(z)=0$. Also, the concavity of $v(z)$ implies $v-v'z>0$.

Thus, the model supplies a theoretical justification for the secular decline of the agricultural sector in the course of economic development, one of the most firmly established empirical generalizations.

With appropriate substitutions and transformations (see Appendix A), the behavior of the economy over time is basically described by the behavior of two variables: per capita income z, and the overall capital intensity measured in terms of the efficiency of industrial labor $\bar{k}$.

$$\frac{Dz}{z} = \frac{1}{M(z)}\left(\alpha + \beta\frac{\lambda}{1-\gamma} - (1-\beta-\delta)\phi(z) + \beta\frac{D\bar{k}}{\bar{k}}\right) \tag{19}$$

$$\frac{D\bar{k}}{\bar{k}} = N(z)\bar{k}^{\gamma-1} - \phi(z) - \frac{\lambda}{1-\gamma} - \eta \tag{20}$$

where

$$M(z) = \frac{v'z}{v} + \left[\delta + \frac{\beta}{1-(1-E)s}\right] \cdot [1-(1-C)s]\left(1 - \frac{v'z}{v}\right) \tag{21}$$

$$N(z) = \frac{\rho[1-(1-C)s]}{[1-(1-E)s]^\gamma} \tag{22}$$

and $M(z)>0$, $N(z)>0$ for all the relevant values of z, $s(z)$, $v(z)$ and $v'(z)$.

III. Stagnation and Steady Development

The system (19) and (20) will terminate its motion at the point $(z^*, \bar{k}^*)$ where $Dz/z=0$ and $D\bar{k}/\bar{k}=0$; at this point

$$\phi(z^*) = \frac{\alpha + \beta\frac{\lambda}{1-\gamma}}{1-\beta-\delta}$$

delineates the maximum sustainable rate of population growth for the model. Sustained long-run development of the economy is impossible unless the actual rate of population growth remains below this figure.

The maximum sustainable rate in this model is determined by the rate of technical

progress in agriculture, the rate of (Harrod-neutral) technical progress in industry (weighted by the production elasticity of capital in agriculture), and the production elasticity of land in agriculture. Thus we see that the introduction of capital into agricultural production will raise the maximum sustainable rate of population growth significantly—by $[\lambda/(1-\gamma)][\beta/(1-\beta-\delta)]$. If the rate of (Harrod-neutral) technical progress in industry is 2 percent, and the production elasticities of capital and land in agriculture are both .2, the maximum sustainable rate will be raised by 2 percent, implying that the economy could achieve steady development with a 2 percent higher rate of population growth. The contribution of the industrial technical progress is greater, the greater the production elasticity of capital in agriculture, or the smaller the production elasticity of land in agriculture. As a special case of the present model, if the production elasticity of capital in agriculture β is zero, the model reproduces the result of conventional models; the maximum sustainable rate is determined as the ratio of the rate of technical progress in agriculture to the production elasticity of land in agriculture.[8]

A. *The Case of Stagnation*

At the point where $Dz/z=0$ and $D\bar{k}/\bar{k}=0$, per capita income ceases to increase any longer. Thus, we call this point $(z^*, \bar{k}^*)$ the point of stagnation. As long as the population growth function $\phi(z)$ is monotonically increasing, the point of stagnation is unique.

The investigation of the behavior of the system around the point of stagnation indicates (see Appendix B) that if

$$\mu \geqq \frac{\xi}{1-\beta} \tag{23}$$

then all movements around the point of stagnation will approach it.[9] Condition (23) is satisfied unless the ratio of the agricultural to the industrial money wage is smaller than the ratio of the relative share of labor in agriculture to the sum of the relative shares of labor and landlord in agriculture. This condition is likely to hold for a wide range of plausible values of the parameters. If there is no wage differential between the two sectors, that is, if $\mu=1$, this condition holds regardless of the values of relative shares. If $\mu=.7$, it holds for $\beta=.2$ and $\xi=.55$. It is more likely to hold the smaller the degree of the wage differential between the two sectors. Alternatively, it is more likely to hold the larger the production elasticity of land in agriculture, since condition (23) can be expressed as $1-\beta-\delta \geqq (\xi-\mu\delta)/\mu$. In terms of capital intensities or labor productivities, this condition is expressed as

$$\frac{k_a}{k_m} \geqq \frac{(1-\gamma)/\gamma}{(1-\beta)/\beta} \quad \text{or} \quad \frac{y_a}{y_m/q} \geqq \frac{1-\gamma}{1-\beta}$$

Even if this condition is not satisfied, the possibility that the economy will approach the point of stagnation is still large.[10] Thus, *as long as the actual growth rate of the population tends to exceed the maximum sustainable rate of population growth, so that the point of stagnation exists, the economy most likely will approach the point of stagnation. A rise in the savings ratio does not play any role in assisting the economy to escape from the trap.*

[8] See Sato and the author.

[9] The phase diagram for this case would indicate that as long as the point of stagnation is unique, the economy will approach it from any initial position (see the author (1973a)).

[10] As shown in Appendix B, a sufficient condition for the stability of the stagnation point is $N'(z) \leqq 0$ for which condition (23) is a sufficient condition. Even if condition (23) is not satisfied, $N'(z)$ can still be nonpositive. This possibility is large if the value of s is still large at the stagnation point. Instability of the stagnation point requires

$$N'(z) > \frac{(1-\gamma)MN}{\beta\bar{k}} + \frac{(1-\delta)}{\beta}\phi'(z)\bar{k}^{1-\gamma}$$

which seems difficult to be met by plausible values of parameters and variables. The sign of $N'(z)$ is also related to the form of savings function. If we assume the classical function, i.e., $S=\gamma Y_m$, $N'(z)$ can be shown to be always positive. In general, if we assume $S=\rho_a q Y_a + \rho_m Y_m$ where ρ_a and ρ_m are the savings ratios of the agricultural and industrial sectors, the necessary and sufficient condition for $N'(z) \leqq 0$ is:

$$\frac{\rho_a}{\rho_m} \geqq \frac{1}{C}\left[1 - \frac{\gamma(1-E)}{[1-(1-\gamma)(1-E)s]}\right]$$

Once the point of stagnation is reached, per capita income ceases to increase any longer and the labor force ceases to migrate from agriculture to industry. Agricultural production increases just enough to feed the additional members of the society at the existing level of food consumption; industrial output per capita increases at the rate $\lambda/(1-\gamma)$, but it is accompanied by the same percentage rate of increase in the price of agricultural goods relative to the price of industrial goods. A rise in the savings ratio will raise the trap level of $\bar{k}$ but leave the burgeoning population with the same trap level of per capita income.

B. *The Case of Steady Development*

If the actual growth rate of the population never exceeds the maximum sustainable rate, so that the inequality

$$\phi(z) < \frac{\alpha + \beta \dfrac{\lambda}{1-\gamma}}{1-\beta-\delta}$$

always holds, then the point of stagnation does not exist. In this case the behavior of the economy is characterized by steady development. Per capita income rises continuously and the industrial component of the labor force becomes more and more dominant. In the long run the economy will reach the equilibrium position at which the agricultural and industrial sectors grow in a balanced fashion and per capita income grows at a constant rate, which is similar to the long-run equilibrium position described by neoclassical growth theory.[11]

As a special case of the present model, let us investigate the behavior of an economy for which land is not a limitational factor in agricultural production, so that the agricultural production function exhibits constant returns to scale with respect to capital and labor alone. We can achieve this by ascribing to land the properties of a free good, which does not appear in the production function as a factor limiting output although it may be essential to production.

In this case, the agricultural production function becomes

$$(1') \qquad Y_a = A_0 e^{\alpha t} K_a^{\beta} L_a^{\delta}, \qquad \beta + \delta = 1$$

Since the term $1-\beta-\delta$ becomes zero, the term $(1-\beta-\delta)\phi(z)$ does not occur in the determination of Dz/z, or in equation (19). Thus, *if land is not a limitational factor in agricultural production, the growth of the population never works as a pressure upon the growth of per capita income.* In this case since we have $\alpha+\beta[\lambda/(1-\gamma)]-(1-\beta-\delta)\phi(z)=\alpha+\beta[\lambda/(1-\gamma)]>0$, *regardless of the behavior of the actual growth rate of the population* $\phi(z)$, *there is no stagnation point.* Steady development of the economy will *always* be achieved. In this case the model will reproduce the result of the neoclassical two-sector growth model with different rates of technical progress. The long-run equilibrium is always stable, since the elasticity of substitution is equal to unity in each of the sectors. Thus, the result of the neoclassical two-sector growth model, in which balanced growth is always achieved with Cobb-Douglas type production functions, arises because land does not enter the production function as a factor restricting the level of output.

IV. Concluding Remarks

The results I have obtained seem to elucidate the importance of capitalization of the agricultural sector for countries in the process of development. Introduction of capital into agriculture will greatly enhance

[11] Since the actual growth rate of the population never exceeds the maximum sustainable rate of population growth no matter how large per capita income grows, we may assume that the actual growth rate of the population $\phi(z)$ approaches some constant value n in the long run, which is smaller than the maximum sustainable rate. Then, at the long-run equilibrium, the population grows at the constant percentage rate n, and per capita income grows at the constant positive rate $\alpha+[\beta\lambda/(1-\gamma)]-(1-\beta-\delta)n$. The overall, agricultural and industrial capital intensities all proceed at the rate $\lambda/(1-\gamma)$, so that the overall capital intensity in efficiency units $\bar{k}$ stabilizes at a constant value. Also, under the condition $\lim_{z\to\infty} v'(z)=\epsilon$, $0<\epsilon<1$, which states that the fraction of each additional unit of income spent on agricultural commodities will never be zero, no matter how large per capita income may grow, the agricultural component of the total labor force stabilizes at a constant value.

the possibility of steady development and provide policy makers with a new useful policy measure to escape the trap of stagnation. In addition, the introduction of capital into agriculture is likely to infuse modern knowledge and spirit into what is frequently a traditionally oriented sector, thereby stimulating improvement in the quality of labor. If technical progress takes place with new capital, or if the increase in the efficiency of labor is accelerated by higher capital intensity (through the learning-by-doing effect), the introduction of capital into the agricultural sector becomes even more important.

Appendix

A. *Derivation of Equations* (18), (19), *and* (20)

We restate the model using the definitions of capital intensities and the proportion of agricultural workers to the total labor force:

Definition of the proportion of agricultural workers to the total labor force:

$$s = \frac{L_a}{L} \tag{A1}$$

Definition of the overall capital intensity:

$$k = \frac{K}{L} \tag{A2}$$

Agriculture production function:

$$y_a = e^{\alpha t} k_a^{\beta} L_a^{\beta+\delta-1} \tag{A3}$$

Industrial production function:

$$y_m = e^{\lambda t} k_m^{\gamma} \tag{A4}$$

Determination of the terms of trade:

$$q = C \frac{y_m}{y_a} \tag{A5}$$

where $C=\mu(1-\gamma)/\xi$.

Distribution of capital between the two sectors:

$$k_a = E k_m \tag{A6}$$

where $E=\mu(1-\gamma)\beta/\xi\gamma$.

$$k = sk_a + (1-s)k_m \tag{A7}$$

Labor supply function:

$$\frac{DL}{L} = \phi(z) \tag{A8}$$

Demand function for agricultural output:

$$v(z) = \theta s y_a \tag{A9}$$

Definition of per capita income:

$$z = \theta\left[s y_a + (1-s)\frac{y_m}{q}\right] \tag{A10}$$

Behavioral equation for capital accumulation:

$$\frac{Dk}{k} = \rho[1-(1-C)s]\frac{y_m}{k} - \phi(z) - \eta \tag{A11}$$

Using equations (A5), (A9), and (A10), we can derive equation (18).

$$s = \frac{v(z)}{v(z) + C(z - v(z))}$$

Substituting equation (A9) into (A3) and using (A1) and (A6) we obtain

$$v(z) = \theta E^{\beta} e^{\alpha t} k_m^{\beta} s^{\beta+\delta} L^{\beta+\delta-1} \tag{A12}$$

Also, substituting equation (A6) into (A7) we obtain

$$k = [1-(1-E)s]k_m \tag{A13}$$

Substituting equation (A4) into (A11) we obtain

$$\frac{Dk}{k} = \rho[1-(1-C)s]e^{\lambda t}\frac{k_m^{\gamma}}{k} - \phi(z) - \eta \tag{A14}$$

Transforming (A12), (A13), (A14), and (18) into growth rates, and eliminating Ds/s and Dk_m/k_m, we obtain

$$\frac{Dz}{z} = \frac{1}{M(z)}\left[\alpha - (1-\beta-\delta)\phi(z) + \beta\frac{Dk}{k}\right] \tag{A15}$$

$$\text{(A16)} \quad \frac{Dk}{k} = N(z)e^{\lambda t}k^{\gamma-1} - \phi(z) - \eta$$

Defining

$$\bar{k} = e^{[\lambda/(\gamma-1)]t}k$$

the above system can be transformed into (19) and (20).

B. *Derivation of Condition* (23)

The behavior of the system around the point of stagnation can be examined by linearizing the system (19) and (20) at the point $z=z^*$ and $\bar{k}=\bar{k}^*$. Let $G(\bar{k}, z)=D\bar{k}/\bar{k}$ and $H(\bar{k}, z)=Dz/z$. Then,

$$G_{\bar{k}} = -(1-\gamma)N(z)\bar{k}^{\gamma-2} < 0$$

$$G_z = \bar{k}^{\gamma-1}N'(z) - \phi'(z)$$

$$H_{\bar{k}} = \frac{\beta G_{\bar{k}}}{M(z)} < 0$$

$$H_z = \frac{1}{M(z)}\left[-(1-\beta-\delta)\phi'(z) + \beta G_z\right]$$

each evaluated at the point $\bar{k}=\bar{k}^*$ and $z=z^*$.

Since $G_{\bar{k}}H_z - G_zH_{\bar{k}} = -(1-\beta-\delta)\phi'(z)\,[G_{\bar{k}}/M(z)]>0$, the product of the two characteristic roots of the system is always positive, as long as $\phi'(z)>0$. Also, if the sign of $N'(z)$ is nonpositive, the sign of G_z will be negative, and then the sum of the two roots will be negative. Since

$$N'(z) = \rho s'(z)\{[\gamma(1-E)-(1-C)](1-s) + [\gamma C(1-E) - E(1-C)]s\}\cdot [1-(1-E)s]^{-\gamma-1}$$

where $s'(z)<0$, if the term

$$J = [\gamma(1-E)-(1-C)](1-s) + [\gamma C(1-E) - E(1-C)]s$$

is nonnegative, then $N'(z)\leqq 0$. However, both

$$[\gamma(1-E)-(1-C)] = \frac{(1-\gamma)(1-\beta)}{\xi}\left(\mu - \frac{\xi}{1-\beta}\right)$$

and

$$[\gamma C(1-E) - E(1-C)] = \frac{C}{\gamma\xi}\left[(\mu\beta+\xi)\gamma^2 - 2\mu\beta\gamma + \beta(\mu-\xi)\right]$$

can be shown to be nonnegative under condition (23).

REFERENCES

J. C. H. Fei and G. Ranis, *Development of the Labor Surplus Economy: Theory and Policy*, Homewood 1964.

B. F. Johnston, "Agriculture and Structural Transformation in Developing Countries: A Survey of Research," *J. Econ. Lit.*, June 1970, *8*, 369–404.

D. W. Jorgenson, "The Development of a Dual Economy," *Econ. J.*, June 1961, *71*, 309–34.

———, "Surplus Agricultural Labour and the Development of a Dual Economy," *Oxford Econ. Pap.*, Nov. 1967, *19*, 288–312.

A. C. Kelley, J. G. Williamson, and R. J. Cheetham, *Dualistic Economic Development*, Chicago 1972.

H. Leibenstein, *Economic Backwardness and Economic Growth*, New York 1957.

W. A. Lewis, "Economic Development with Unlimited Supplies of Labour," *Manchester Sch. Econ. Soc. Stud.*, May 1954, *22*, 139–91.

Y. Niho, (1973a) "Population Growth and the Development of a Dual Economy: Analysis and Some Testing," unpublished doctoral dissertation, Brown Univ. 1973.

———, (1973b) "Population Growth, Agricultural Capital and the Development of a Dual Economy: An Application to Germany and the United States," paper presented at the 1973 Winter Meetings of the Econometric Society, New York.

R. Sato, *The Theory of Economic Growth*, Tokyo 1968.

———, "Stability Conditions in Two-Sector Models of Economic Growth," *J. Econ. Theory*, June 1969, *1*, 107–17.

——— and Y. Niho, "Population Growth and the Development of a Dual Economy," *Oxford Econ. Pap.*, Nov. 1971, *23*, 418–36.

P. Zarembka, "Marketable Surplus and Growth in the Dual Economy," *J. Econ. Theory*, June 1970, *2*, 107–21.

Comment on Corporate Income Taxes and the Cost of Capital: A Correction

By Pao L. Cheng*

In a communication in this *Review*, Franco Modigliani and Merton Miller interpreted their formula for the variance of the after-tax returns of a levered firm in risk-equivalent class k:

$$(1)\quad \text{Var}\,(X^\tau) = \sigma^2(\overline{X}^\tau)^2[1 - \tau(R/\overline{X}^\tau)]^2$$

where

$X^\tau = (1-\tau)(X-R)+R$, X being the earnings before interest and taxes

$R = rD$, the default-free interest bill

$\sigma^2 = \text{Var}(Z)$, with Z defined as $X/\overline{X}$

As follows:

> Thus, if the tax rate is other than zero, the shape of the distribution of X^τ will depend not only on the "scale" of the stream $\overline{X}^\tau$ and on the distribution of Z, but also on the tax rate and the degree of leverage (one measure of which is $R/\overline{X}^\tau$). [p. 435]

Equation (1) appears to imply that given a marginal corporate income tax rate τ (assumed equal to the average), the variance of X^τ is smaller the higher the degree of leverage, since the interest bill R is positively related to D. I shall show, however, that Var (X^τ) is independent of the degree of leverage and the scale of the stream $\overline{X}^\tau$. The demonstration turns out to be surprisingly simple:

$$\begin{aligned}\text{Var}\,(X^\tau) &= \sigma^2(\overline{X}^\tau)^2[1 - \tau(R/\overline{X}^\tau)]^2\\ &= \sigma^2(\overline{X}^\tau - \tau R)^2\\ &= \sigma^2[(1-\tau)(\overline{X} - R) + R - \tau R]^2\\ &= \sigma^2(1-\tau)^2\overline{X}^2\end{aligned}$$

Hence, the variance of after-tax return streams depends only on the distribution of Z, the corporate income tax rate τ, and the mean of X.

Alternatively, we could go directly to Modigliani and Miller's equation (1), p. 435, where X^τ was defined either as a function of the random variable X or Z:

$$\begin{aligned}X^\tau &= (1-\tau)(X-R) + R\\ &= (1-\tau)X + \tau R\\ &= (1-\tau)\overline{X}Z + \tau R\end{aligned}$$

By inspection we see immediately that

$$\begin{aligned}(2)\qquad \text{Var}\,(X^\tau) &= (1-\tau)^2\,\text{Var}\,(X)\\ &= (1-\tau)^2\overline{X}^2\sigma^2\end{aligned}$$

where τR, the tax savings on interest bills, is a constant.

Moreover, in their footnote 4, Modigliani and Miller asserted:

> It may seem paradoxical at first to say that leverage *reduces* the variability of outcomes, but remember we are here discussing the variability of total returns, interest plus net profits. The variability of stockholder net profits will, of course, be greater in the presence than in the absence of leverage, though relatively less so than in an otherwise comparable world of no taxes. [p. 435]

Obviously, this assertion is again incorrect. Defining stockholder net profits as

$$\begin{aligned}\pi^\tau &= (1-\tau)(X-R)\\ &= (1-\tau)X - (1-\tau)R\\ &= (1-\tau)\overline{X}Z - (1-\tau)R\end{aligned}$$

we see immediately that

$$\begin{aligned}(3)\qquad \text{Var}\,(\pi^\tau) &= (1-\tau)^2\,\text{Var}\,(X)\\ &= (1-\tau)^2\overline{X}^2\sigma^2\end{aligned}$$

Equations (2) and (3) show that Var (π^τ)

* Professor of finance, University of Massachusetts and Simon Fraser University.

and Var (X^{τ}) are identical, both independent of the degree of leverage. Hence, there is no paradox. However, by virtue of the relationship $X^{\tau}=\pi^{\tau}+R$, any nonintuitive feeling about the fact that residual return π^{τ} for stockholders and total return X^{τ} are identical, except for their location, cannot endure.

REFERENCE

F. Modigliani and M. Miller, "Corporate Income Taxes and the Cost of Capital: A Correction," *Amer. Econ. Rev.*, June 1963, *53*, 433–43.

Capital and Technology Movements and Economic Welfare

By NOBUO MINABE*

Economists have long valued capital mobility as an aid to international efficiency. Robert Mundell has shown that, under certain conditions, the capital movement induced by tariffs on goods will substitute for trade and enable the tariff-impeded international economy to reach essentially the same equilibrium as would a completely free international economy. It will also (except possibly for the case when the optimal tariff is non-zero) benefit both countries. But if perfect, unlimited capital mobility is a good thing for the borrowing country, it does not follow, as this paper demonstrates, that any limited amount of capital mobility, however small, is a good thing. If the capital movement is due to discrepancies in returns due solely to a tariff, then a small amount of capital inflow will actually diminish welfare in the recipient country. Thus, countries which take a few cautious steps toward liberalization of capital flow may find their welfare diminished. This conclusion is related to the intuitive notion that perhaps it might not be a good idea to allow only a few foreign investors to reap high rates of return in capital-short countries. If foreign investment is allowed, enough should be allowed to depress the rate of return so that "excessive" profits are not taken out of the country by a few privileged investors. On the other hand, however, this paper shows that if differences in returns are due to a technological advantage in the capital-importing country so that the country with initially high return to capital nonetheless exports the capital-intensive good, and if these differences in returns are accompanied but not caused by tariffs, then even short, hesitant steps toward liberalization of capital inflow will improve the borrowing country's welfare. This paper also shows that the home country's welfare may be improved or diminished by importing the advanced country's technology when the home country levies a tariff and pays royalties for the technology. The critical point is that a movement of technology may either increase or diminish economic welfare, depending on whether the technological progress is of the capital-saving type or of the labor-saving type in the fixed production coefficients case. In the same case, when the technological progress is Hicksian neutral, technology inflow will reduce or have no effect on the home country's economic welfare, depending on whether the foreign country's technological progress occurs in the home country's import-competing or export industries. Similarly, capital or technology inflow will never increase the borrowing country's welfare when the country specializes completely in a single industry. Some of these results should be modified when we treat the possibility that changes in factor prices alter factor proportions.

I. The Patterns of Commodity Trade and Capital Movements

Assume purely competitive market conditions in a country producing two commodities, say manufactured good M and food F, by two factors of production, say labor L and capital K. Entrepreneurs want to minimize the total cost[1]

$$z = wL + rK \tag{1}$$

* Associate professor, Osaka City University. I have benefited from the suggestions and criticism of the managing editor and especially the referees of this paper. William Der has also given useful suggestions. M. Bronfenbrenner and G. Blackford helped to correct the English.

[1] The trade model will be analyzed by using a linear programming model, because my arguments need only some modifications to treat the possibility that changes in factor prices alter factor proportions. The basic model used here is discussed in the author (1974).

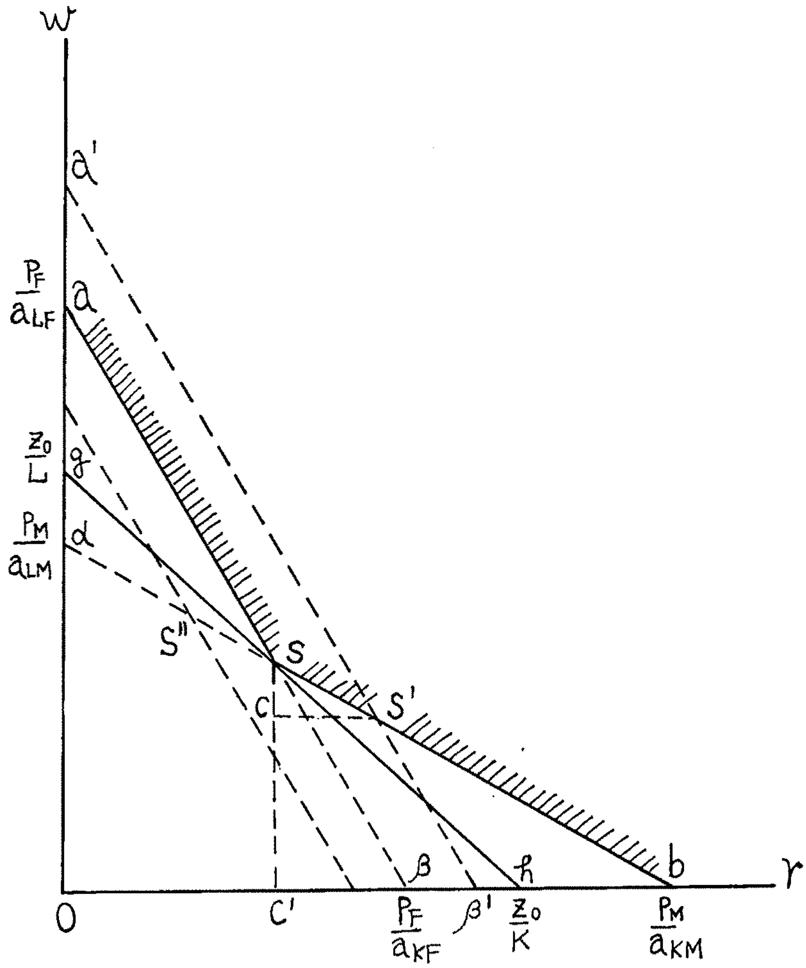


FIGURE 1

subject to

$$(2) \qquad a_{LM}w + a_{KM}r \geq p_M$$
$$a_{LF}w + a_{KF}r \geq p_F$$

and $$w,\ r \geq 0$$

where a_{ij} denotes the quantity of factor i required to produce a unit of commodity j. This is a typical linear programming model, which may be described as the problem of minimizing a single linear relation (1) subject to a number of linear inequalities (2). The solution of the problem will determine factor prices w and r, which will minimize the total cost z without giving rise to excess profits in any possible use of the factor endowments L and K. These data are depicted in Figure 1, where factor prices are measured along the two axes and the lines $a\beta$ and $b\alpha$ represent those combinations of the two factor prices that exactly meet (i.e., with no excess) each of the two requirements in (2). (It is assumed here that M is relatively labor intensive.)[2] The line gh represents equation (1) where z_0 is the solution of the problem. Whenever the commodity prices p_F and p_M are given a priori, the factor prices r and w are uniquely determined at the intersection point s of the two lines $a\beta$ and $b\alpha$ in such a way to minimize

[2] When M is capital intensive, the relationship of the slopes $a\beta$ and $b\alpha$ is reversed.

the total cost so long as the slope of $bs < K/L$ $<$the slope of as.[3] This is the well-known Heckscher-Ohlin-Samuelson factor price equalization theorem. That is, factor prices r and w are equalized everywhere in the competitive world (commodity) market. Therefore, if no country specializes in a single commodity and if production functions are common to all countries, then there are no capital movements because the return to capital r is the same everywhere. On the other hand, 1) when a country levies a tariff on goods, or 2) when production functions differ between countries, or 3) when a country specializes completely in a single industry, then capital will move internationally in response to differentials in the return to capital r.

Let us consider case 1) first. Suppose a country levies an import tariff on F so that p_F rises in the domestic market.[4] (Throughout most of this paper, it is assumed without loss of generality that the country imports food.) A rise in p_F will shift the $a\beta$ line to the $a'\beta'$ line, which in turn will shift the intersection point from s to s'. The value of r rises more than proportionally, and w falls because $\Delta r/r = cs'/0c' > \beta\beta'/0\beta = \Delta p_F/p_F$ and $\Delta w/w = -(sc/c's)$.[5]

An example of the second case is when Hicksian-neutral technological progress occurs in industry F which is capital intensive. Then the effect on factor prices is the same as in the previous case assuming constant commodity prices. When commodity prices are held constant, a proportionate decrease in a_{LF} and a_{KF} (as reflecting the neutral technological progress) will also bring about a shift of the intersections of the two lines from s to s'. Then as in case 1), r rises more than proportionally compared to the degree of technological progress, and w falls assuming constant commodity prices.

We now consider case 3). When the factor endowment ratio K/L happens to be equal to or greater than the slope of as (or alternatively, equal to or less than the slope bs), this would imply actual losses in M, and the economy specializes completely in industry F (or in the alternative case M). In these cases, the factor prices can be anywhere or the intercept of one (or the other) of the two lines as (or bs) at given commodity prices p_F and p_M. Thus, whenever a country specializes completely in a single industry, factor prices are not equalized, so that there is incentive for capital to move internationally.

Notice that in case 1) the import tariff induces owners of capital to send it to the home country (which has imposed the tariff) if, and only if, the home country is an importer of the capital-intensive good F. (Conversely, if the home country is an importer of the labor-intensive good M, owners of capital in that country will send it to the foreign country.)[6]

However, this Heckscher-Ohlin capital movement pattern is no longer true in case 2). Hicksian-neutral technological progress in one (say, F) industry in one country only may bring about a rise in r and also decrease the commodity price p_F in that industry.[7] Thus a country which has technological progress in the capital-intensive good industry may have a comparative advantage in the capital-intensive good, but nonetheless the country may be a capital importer![8] However, if p_F falls sufficiently so that the $a\beta$

[3] This is a sufficient condition for incomplete specialization.

[4] When a country is small compared to the rest of the world, Lloyd Metzler's effect of tariffs on domestic prices is always excluded.

[5] The same thing is true in the case of nonfixed production coefficients. This is the well-known Stopler-Samuelson tariff theorem.

[6] According to the Heckscher-Ohlin theory of trade patterns, an importer of the capital-intensive good is considered a capital-scarce country. Therefore, if foreign owners of capital send it to this country, we may define it as the Heckscher-Ohlin pattern of capital movements because a capital scarce county imports capital.

[7] The relationship between technological progress and the pattern of trade has been discussed by the author (1966).

[8] When a country exports a capital-intensive good, the country may be defined a capital-abundant country. The Leontief paradox type of capital movement occurs if the country is a capital importer, because a capital-abundant country imports capital. If the production functions are the same in all countries, the Leontief paradox pattern of capital movements may occur when there exist factor intensity reversals in the case of neoclassical nonfixed production coefficients.

line shifts downwards until the intersection of the two lines in Figure 1 moves to a point such as s'', then capital moves out of the country which is exporting the capital-intensive good (or the country which is importing the capital-intensive good will also be an importer of capital).

II. Capital Movements and Economic Welfare: The Case of Incomplete Specialization

The dual of our trade model can be written as follows: Entrepreneurs want to maximize total revenue

$$Z = p_M M + p_F F \tag{3}$$

subject to the factor endowments constraint

$$a_{LM}M + a_{LF}F \leq L \tag{4}$$

$$a_{KM}M + a_{KF}F \leq K$$

and

$$F, M \geq 0$$

where F and M represent the outputs of the two commodities, and p_F and p_M show the commodity prices in the domestic market in this case. In Figure 2 the line GH represents equation (3), where Z_0 is the solution of the problem at a given domestic commodity price ratio p_F/p_M with slope of GH. When manufactured goods are relatively labor intensive, the lines AE and BD represent the labor and capital budget equations, respectively. (If manufactured goods M are relatively capital intensive, the relationship between the slopes of AE and BD must be reversed.) The locus ASB represents the "efficiency frontier" of the economy, or the pro-

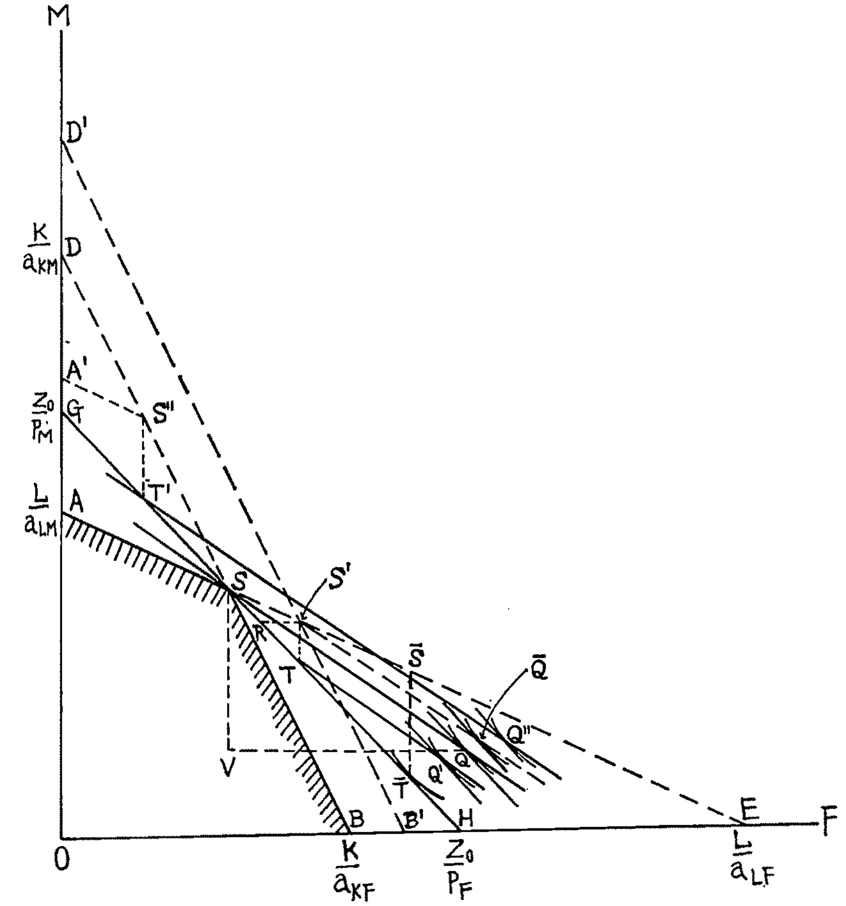


FIGURE 2

duction possibility curve. When a domestic commodity price ratio p_F/p_M is given, the outputs of the two commodities are uniquely determined at the corner point S in such a way as to maximize Z so long as the slope of $AS < p_F/p_M <$ the slope of BS.[9]

We assume the country to be small and so it faces given terms of trade represented by the line SQ. If the government of the economy is levying an import tariff on F, the inverse of the slope of SQ would represent the international commodity price ratio $p_F/p_M \cdot(1+t)$, which differs from the domestic price ratio of the slope of GH to an extent determined by the rate t of protection of F. In the world market, the line SQ shows the home country's budget line so that Q represents the consumption (equilibrium) point, and welfare is maximized at that point facing a given domestic commodity price ratio of the slope of GH. On the other hand, the production (equilibrium) point will be at S at the same domestic commodity price ratio. The country is importing VQ of F in exchange for SV of M at the equilibrium terms of trade SQ in the world market.

Let us consider the effect of capital movements on the home country's economic welfare (or real income). We now presume the country to be a capital importer and to levy an import duty on F. Our country is so small relative to the rest of the world that its production conditions and factor endowments can have no effect on the world commodity prices. In the following analysis it is assumed that both goods are normal goods (and the existence of inferior goods is excluded).

First we analyze the case in which manufactured goods are labor intensive and food is capital intensive. Capital inflow (which means an increase of K) will shift the corner point of line DB from D to D', and the new production equilibrium point will be at S'.[10] The increase of the net value product represents the marginal productivity of the foreign capital, which must eventually be repatriated.[11] This total return to foreign capital is RS' measured in units of F, or $S'T$ measured in units of M at a given domestic commodity price ratio GH. The foreign country benefits more by receiving the return to capital r in M, because its domestic price of M is lower than its international price. Thus the home country may have to pay the return to capital r in M.[12] The new utility level of the home country is given by the community indifference curve which intersects a new international price line TQ' (parallel to SQ) through point T with a slope equal to that of GH.[13] Therefore the home country is made worse off by the small movement of capital.[14]

Now suppose that manufactured goods are capital intensive and food is labor intensive. In this case the relationship of the slopes of AE and BD should be reversed in Figure 2. Capital inflow will shift the corner point from A to A', and the equilibrium production point will be at S''. The new utility level of the country is given by the community indifference curve that intersects a

[9] The commodity price ratio must be in a certain range for the country to produce both goods.

[10] The output of F increases more than proportionally, and that of M decreases at constant domestic commodity prices. This is the well-known Rybczynski effect.

[11] I am indebted for the following proof of this statement to a commentator on an earlier draft of the paper. We have the following relationship, by Euler's theorem:

$$p_1X_1 = wL_1 + rK_1, \qquad p_2X_2 = wL_2 + rK_2$$

where X_i, L_i, K_i are outputs and factors in industry i.

$$\begin{aligned} p_1X_1 + p_2X_2 &= wL_1 + rK_1 + wL_2 + rK_2 \\ \Delta(p_1X_1 + p_2X_2) &= p_1\Delta X_1 + p_2\Delta X_2 \\ &= \Delta(wL_1 + rK_1 + wL_2 + rK_2) \\ &= r\Delta K_1 + r\Delta K_2 = r(\Delta K_1 + \Delta K_2) \\ &= r(\text{capital inflow}) \end{aligned}$$

since p_1, p_2, r, w, and (L_1+L_2) are constant.

[12] It is clear from Figure 2 that the home country is made worse off even if the foreign country receives its rental payments in F, the imported good. But my point here is that if the foreign country receives the return to capital r in F, the welfare loss from capital movements will be reduced in the borrowing country (and the welfare gain will be reduced in the foreign country).

[13] Notice that if the economic expansion is due to the own country's capital accumulation instead of capital import, the home country is made better off, because the new utility level of the home country is given by the community indifference curve which intersects the international price line $S'\bar{Q}$ (parallel to SQ) through point S' with a domestic commodity price of a slope of GH. However, this is not always true (as shown in fn. 23). Capital accumulation may reduce the home country's welfare when the country levies a tariff on goods.

[14] This possibility was pointed out by Hirofumi Uzawa.

new international price line $T'Q''$ (parallel to SQ) through point T' with a slope equal to that of GH, where $S''T'$ is the total return to foreign capital.[15] It is obvious that the country is made better off by the small movement of capital.[16]

Thus it has been shown that a small movement of capital may either decrease or increase economic welfare, depending on whether the country with initially high return to capital is an importer of the capital-intensive good, or an exporter of the capital-intensive good. More specifically:

1) When the returns to capital vary only because of a tariff, a limited movement of capital will make the capital importing country worse off.

2) When the returns to capital vary because of different technologies, and when in addition there is a tariff, a small movement of capital may make the capital importing country either better or worse off.

In reference to argument 1), Mundell asserts (using the same model) that the capital movement caused by discrepancies in returns due to a tariff on goods will achieve the same equilibrium as the free trade equilibrium. This may be the reason why economists have long valued capital mobility as an aid to international efficiency. We should notice here that we are not comparing equilibrium positions. Capital returns are just as unequal after a small amount of capital moves as before. It seems quite realistic that countries often allow in some capital, but not enough to equalize returns (even accounting for risk differentials), and it is very useful to look at nonequilibrium positions. Since the international price and the tariff level are given, as long as there is any trade the domestic price is given as the foreign price plus the tariff. The return to capital is determined by the domestic price ratio, and remains higher than the return to capital abroad. If enough capital moves so that the production point shifts from S to $\overline{S}$ in Figure 3, the domestic price is still equal to the foreign price plus the tariff, where $\overline{ST}$ is the return to foreign capital (see also Figure 2). But whenever owners of capital send the home country more than this amount, the domestic price becomes lower than the foreign price plus the tariff, and trade ceases. However, capital will still continue to move so long as the domestic (commodity) price is higher than the foreign price, because the return to capital is still higher than the return to capital abroad. Since there is no trade and the link between the foreign price and the domestic price is broken, the domestic price can fall, and the return to capital begins to be depressed. Payments to foreign capital depress welfare less and less. Eventually capital moves in sufficiently so that the production point shifts to $\overline{S}'$, which causes the domestic price to fall to the international price and the tariff is completely inoperative, no longer serving to distort the domestic price. At the final domestic equilibrium point q, $\overline{S}'q$ is repatriated to foreign capital, and factor prices are equalized between the two countries. This is nothing but the equilibrium situation Mundell has suggested. Therefore, if foreign investment is allowed in this case, enough should be allowed to depress the rate of return so that "excessive" profits are not taken out of the country by a few privileged investors.

Somewhat similar remarks apply to argument 2). When capital inflow results in an expansion of exports, there is no question of trade being displaced. In this case, given a fixed international price, the return to capital stays higher until complete specialization ensues.[17] Welfare increases somewhat from the beginning as capital inflows, by inducing trade, counteract the allocative distortions of the tariff. When capital inflow shifts the

[15] In this case, if the economic expansion is due to capital accumulation instead of capital inflow, the home country must be made better off.

[16] Might not a shortage of domestic capital, or a strong demand, result in the home country importing good M, which is capital intensive, even though it has a technological advantage in producing it? If there is a tariff, a small amount of capital inflow depresses welfare in such a case. However, this case reduces theoretically to that outlined above in which the home country, which is a capital importer, imports the capital-intensive good.

[17] The situation of complete specialization will be discussed in the next section.

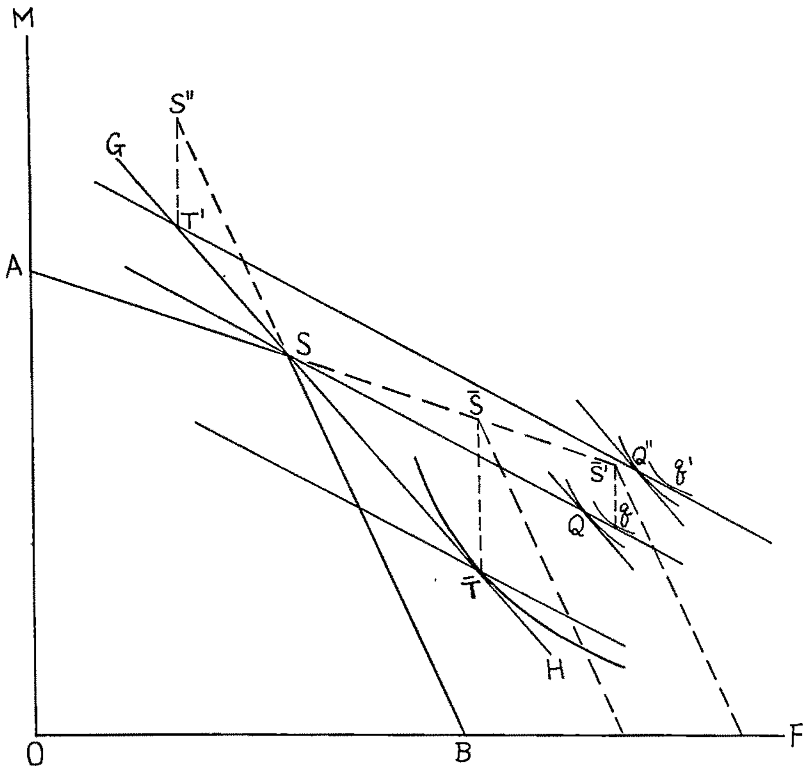


FIGURE 3

production point from S to S'' in Figure 3, the domestic consumption point may be represented by Q''. This tariff-impeded consumption equilibrium with capital mobility may make the capital importing country better off, to an even greater extent than the free trade equilibrium (without capital mobility) represented by point q. This may show the possibility of net gain from capital movements. Of course, the tariff-impeded trade equilibrium without capital mobility represented by point Q will reduce welfare compared to the free trade equilibrium (represented by point q). The same thing is true if capital movement is allowed, that is, the free trade equilibrium with capital mobility represented by point q' is better than the tariff-impeded trade equilibrium with capital mobility represented by point Q''. In equilibrium, with or without capital mobility, the free trade situation is above the tariff situation.

III. Capital Movements in the Case of Complete Specialization

We continue to assume that the home country is a capital importer and levies an import duty. Food may be, relatively, either capital intensive or labor intensive. First consider the case in which F is capital intensive. The country specializes completely in F[18] if the capital-labor endowment ratio K/L

[18] Might not the economy specialize completely in F if the import tariff on F is sufficiently high so that the domestic commodity price ratio $p_F/p_M >$ the slope of BS in Figure 2? Specialization in F would require exportation of F in order for M to be available for domestic consumption (that is, the pattern of trade will be reversed). However, the assumption of perfect competition and the presence of the tariff forbids export of F. If the domestic price of F is p_F and the world price is $p_F/(1+t)$, where t is the tariff rate, then the sale of F abroad would result in a loss to domestic producers since they would only get $p_F/(1+t)$ for each unit on the world market while the production cost of each unit is p_F by the assumption of perfect competition. An im-

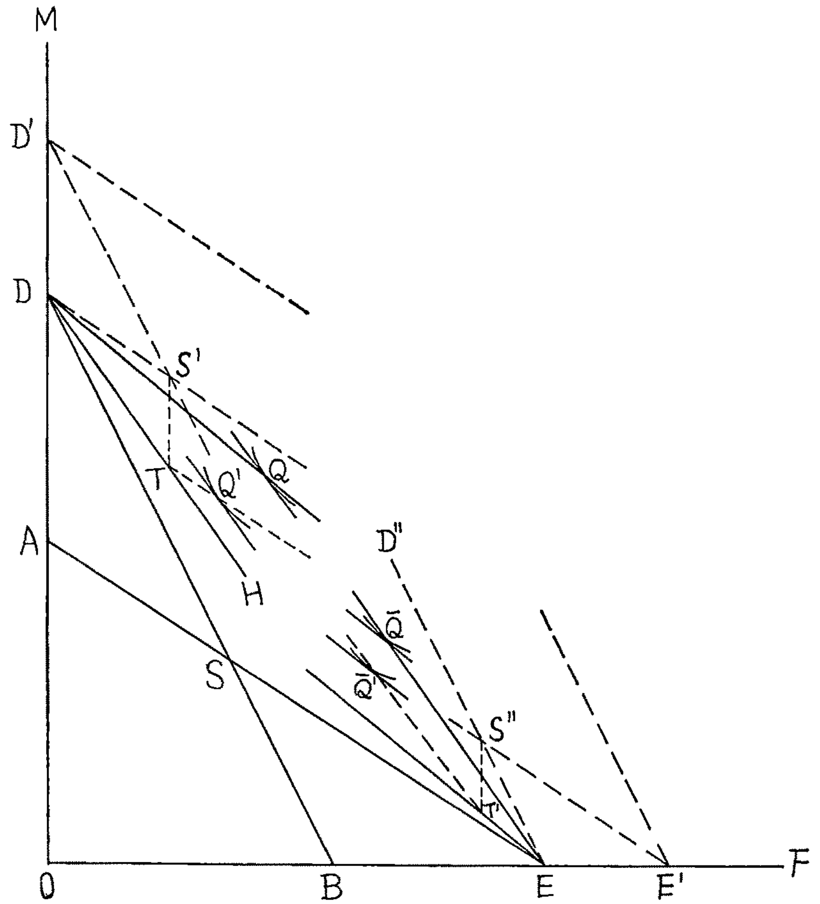


FIGURE 4

happens to be equal to the slope of line sa in Figure 1, or $(K/L)=(a_{KF}/a_{LF})$, that is, $(L/a_{LF})=(K/a_{KF})$.[19] Thus this would correspond, as shown in Figure 4, to changing factor endowments so that the horizontal intercepts of the factor constraint lines become the same (for example, at E), with production taking place at that point. The efficiency frontier becomes a triangle OAE. Capital inflow, which shifts the capital constraint line ED'' from E to E' in Figure 4, suddenly depresses the value of the marginal product of capital to zero, because an increase in K (as reflecting capital inflow) has no effect on both M and F, and capital becomes a free good. Therefore, there is no incentive for capital to move.

port duty which is sufficiently high would eliminate importation of F. Any tariff rate beyond the minimum rate necessary to achieve this elimination is redundant. At the minimum rate, the home country becomes self-sufficient. At this rate, the home country consumes and produces at the point where the highest indifference curve is tangent to the production possibility curve with the domestic price ratio indicated by the slope of the indifference curve at the tangency point. Any further rise in the tariff beyond the minimum rate will not alter the domestic price ratio. For if the domestic price of F were to rise further with a further rise in the tariff rate, more F would be produced while less F would be consumed domestically, causing a domestic surplus of F. Since producers cannot sell F abroad except at a loss, there is pressure to reduce the price of F. (I am indebted for this note to William Der.)

[19] We exclude the case in which the country specializes completely in F when $K/L>$ the slope of as in Figure 1. In this case capital becomes a free good, so that foreign owners of capital will never send it to this country.

On the other hand, when the capital-labor endowment ratio happens to be equal to the slope of line sb in Figure 1, or $(K/L) = (a_{KM}/a_{LM})$, that is, $K/a_{KM}=L/a_{LM}$, the economy specializes completely in M. Thus, the labor endowment constraint in (4) will be ineffective, and the corresponding production possibility curve becomes a triangle OBD as shown in Figure 4. The production point is represented by point D at a domestic commodity price ratio of the slope of DH, while consumption will be at point Q at a given international commodity price ratio of the slope of DQ. Capital inflow (which shifts the line DS upward from D to D') shifts the new equilibrium production point S' to the southeast of the initial production point D. Now the country begins to produce both commodities. The new utility level of the country is given by the community indifference curve that intersects a new international price line TQ' (parallel to DQ) through point T with a slope equal to that of DH (which is a domestic commodity price ratio), where $S'T$ is the total return to foreign capital. The country is made worse off by the small movement of capital.[20]

Now, suppose F is relatively labor intensive. The country specializes completely in M if K/L happens to be equal to the production coefficient ratio in M, a_{KM}/a_{LM}. The corresponding production possibility curve becomes a triangle OBD, and the capital inflow (which shifts the capital constraint line DS' from D to D') has no effect on both F and M in this case, because capital suddenly becomes a free good.[21] Thus, the capital inflow has no effect on the home country's economic welfare. On the other hand, if the factor endowment ratio K/L happens to be equal to the production coefficient ratio in F, a_{KF}/a_{LF}, the corresponding production possibility curve becomes a triangle OAE, and the country specializes completely in F.[22] Capital inflow (which corresponds to having the AE line shift upward from E to E') will shift the production point from E to S'' at given domestic and international commodity price ratios of the slopes of ET' and $E\overline{Q}$. The domestic consumption point will shift from $\overline{Q}$ to $\overline{Q}'$, and the country must be made worse off. Therefore, we may conclude that when a country specializes completely in a single commodity and levies an import duty, a small movement of capital will never improve the borrowing country's economic welfare.

Some of our discussions should be modified if we consider the possibility of substitution of the two factors, labor and capital. Whenever the country specializes completely in the capital-intensive good, capital inflow breaks the link between prices and returns to capital. As long as the return to capital is higher at home than abroad, capital inflow will be continued, but the return to foreign capital is depressed (but does not suddenly become zero as in the fixed-coefficient case) and intramarginal units are paid less than their contribution to outputs. In this case welfare can increase somewhat from the beginning as capital flows in if the country specializes completely in the capital-intensive good. However, welfare must fall when the country specializes completely in the labor-intensive good.

IV. Technology Movements and Economic Welfare

In the traditional Heckscher-Ohlin trade theory, it is assumed that technology moves internationally without any costs. However this may not be true: there are technological differences between countries, and a developing country may import technology from an advanced country and may have to pay royalties for it. The crucial difference between capital and technology movements is that owners of advanced technology obtain its marginal value products by sending the

[20] These discussions are true if the capital-labor endowment ratio K/L is less than the slope of line sb in Figure 1 as long as the labor constraint line DS' does not shift up above the point D' in Figure 4.

[21] As in fn. 18 we exclude the case in which the country specializes completely in M because of a sufficiently high import tariff on M.

[22] When the commodity price ratio p_F/p_M is sufficiently high, the country specializes completely in F. But we exclude this case because the return to capital r becomes zero, so that foreign owners of capital will never send it to this country.

technology abroad but they continue to use the same technology in the home country too; while in the case of capital movements the quantity of capital endowment decreases in the home country after the owners of capital send it to the foreign country. Therefore, capital moves internationally in response to (absolute) differentials in the return to capital, but the owners of technology send it to the foreign country because of the *additional* return to technology (and may even control the local entrepreneurs employing the local capital and labor supplies). That is, as shown below, technology may move from a country whose return to capital is lower to a country in which the return to capital is higher. (Notice that if technological conditions become the same through the technology movement and there are no factor intensity reversals between countries, factor prices are equalized so that only the Heckscher-Ohlin pattern of capital movements occurs when capital moves internationally.)

We assume here without loss of generality that manufactured goods M are relatively labor intensive. Suppose that capital-saving technological progress in F which decreases the value of a_{KF} and shifts the line $a\beta$ to $a\beta'$ in Figure 5 occurs in the foreign country. The equilibrium point will shift from s to s'. Then if owners of the technology send it to the home country, r rises (from r_1 to r_2) and w falls (from w_1 to w_2) in the home country, assuming constant domestic commodity prices p_F and p_M. That is, technology inflow will change the income distribution.

Now let us consider the effect of technol-

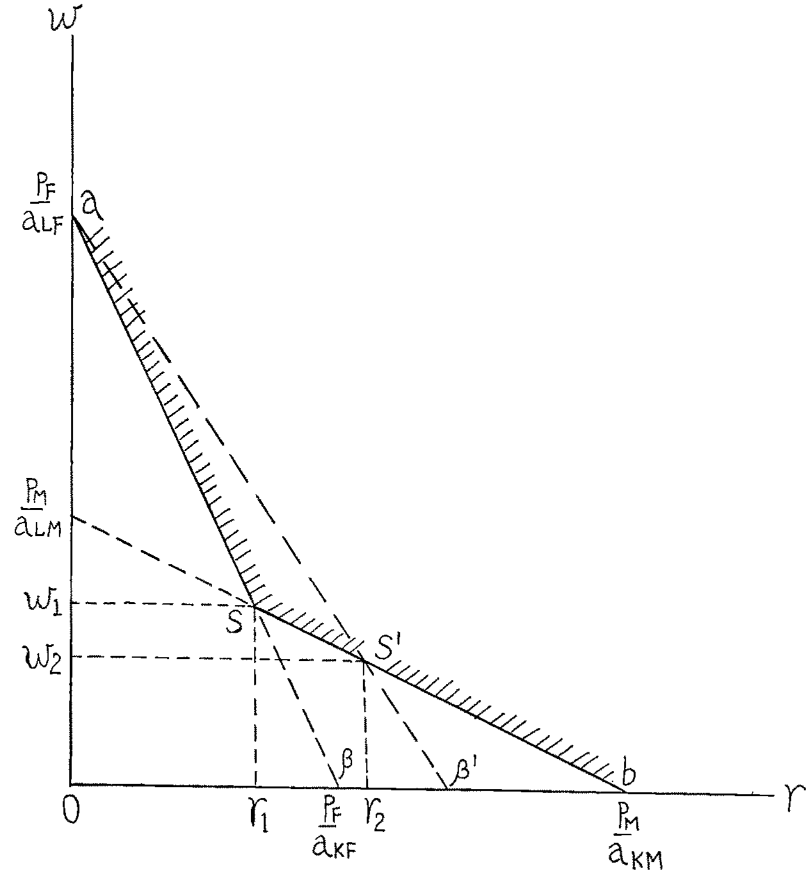


FIGURE 5

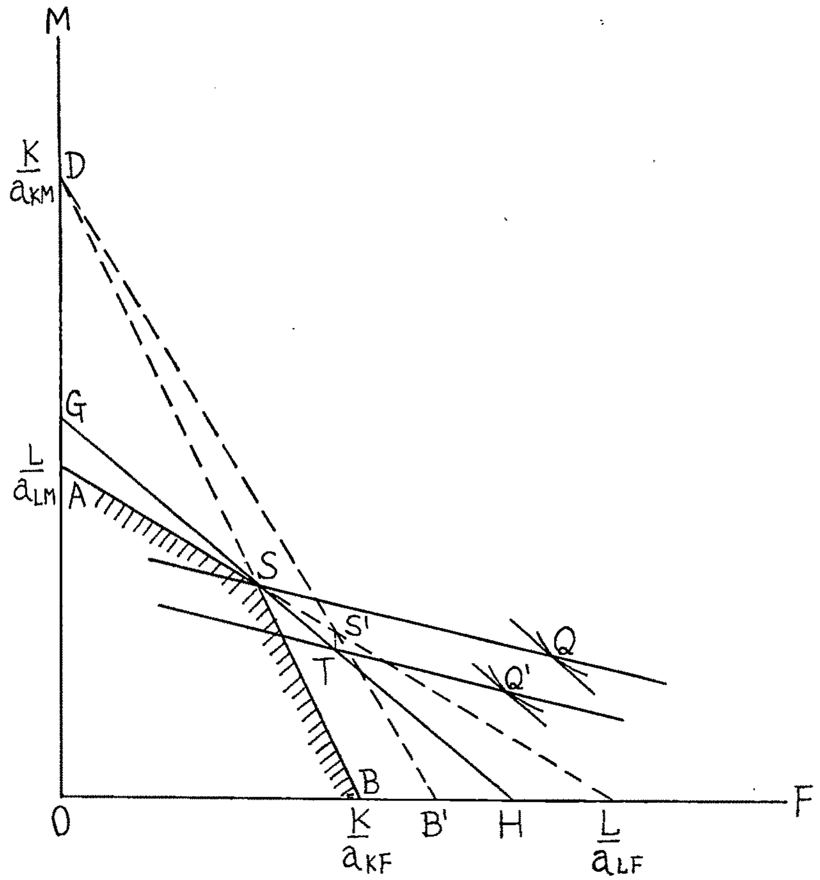


FIGURE 6

ogy movement on economic welfare. In Figure 6 an import of technology of the capital-saving type in F (which means a decrease of a_{KF}) will shift the capital budget line DB to DB', and the new production point will shift from S to S'.[23] The increase of net value product represents the marginal productivity of the foreign technology, which must eventually be repatriated as royalties. This is $S'T$ measured in units of M.

[23] As shown in Figure 6, when a country is following a protective policy, improved efficiency in the protected industry (here, industry F) or accumulation of the factor used intensively (here, capital) in that industry, will not necessarily increase the country's welfare. This fact has been pointed out by Harry Johnson. However, our arguments go beyond it in a significant way by attributing technological progress (or capital increase) to foreign investment, with concomitant complication of equilibrium amounts of movements and repatriated earnings.

When the home country levies an import tariff, the equilibrium consumption point will shift from Q to Q'. That is, the new utility level of the home country is given by the community indifference curve which intersects a new budget line TQ' (parallel to SQ) through point T with a slope equal to that of GH. The country is made worse off by the movement of technology. In this case, the movement of technology raises r and decreases w, but it decreases the home country's economic welfare.

World enterprises quite often invest their technology abroad. They usually do not send their capital abroad, but employ local capital and labor supplies. When a country levies a tariff on goods and pays royalties for importing a foreign country's technology, such a technology inflow will change the income dis-

TABLE 1

Kinds of Technological Progress	Changes of r and w	Economic Welfare
Capital-saving in F	r increases, w decreases	decreases
Capital-saving in M	r decreases, w increases	decreases
Labor-saving in F	r increases, w decreases	increases
Labor-saving in M	r decreases, w increases	increases
Neutral in F	r increases, w decreases	decreases
Neutral in M	r decreases, w increases	unchanged

tribution in the borrowing country, and may decrease (increase) the borrowing country's economic welfare. The direction of changes depends upon the nature of technological progress in an industry which will be invested abroad in an advanced country. Table 1 summarizes these results in the fixed production coefficient case.

When changes in factor prices alter factor proportions, technology inflow will improve the home country's economic welfare when the technological progress is of the Hicksian-neutral type in M. In the non-fixed-production coefficient case, technology inflow may either reduce or improve the home country's economic welfare when the import-competing industry F (which is capital intensive) uses foreign technology of the labor-saving type, or when the export industry M (which is labor intensive) uses foreign technology of the capital-saving type. All other results are the same in the non-fixed-production coefficient case as in the fixed-production coefficient case.

It may be interesting to note that the effects of capital and technology movements on factor prices and economic welfare are analyzed by the same method. We can easily prove that when a country specializes completely in a single industry and levies an import duty, technology inflow will never improve the home country's economic welfare in the fixed-production coefficient case. As in the case of capital movements, some of these conclusions should be modified if the two factors are substitutable. For instance, when the economy specializes completely in the capital-intensive good, the import of technology of the capital-saving type in the same industry reduces (but not to zero as in the fixed-production coefficient case) the return to foreign capital, and can improve the borrowing country's economic welfare.

V. The Core of Our Results

In our discussion, since the home country by assumption is a small one, there is no terms of trade effect, either for capital or for technology. Nor is there a tax on capital or technology flows, so that the extra value of production at home (stimulated by the capital or technology flow) eventually goes abroad to foreign capitalists. That leaves only one way in which welfare can be affected: because of the tariff, any increase in imports stimulated by the capital and the technology flows will raise welfare at home since the cost of obtaining these imports (evaluated at the world price) is less than their worth at home (represented by the home price gross of tariff). Also, imports increase if, and only if, the flow of capital or technology has shifted resources out of import-competing goods into exports. This happens if, and only if, these exports are capital intensive (despite the country being a capital importer) as in the case of capital movement, or if, and only if, technological progress is labor saving in either industry in the case of technology movement. On the other hand, if capital or technology inflow shifts resources out of exports into import-competing goods, the capital or the technology inflows will reduce the home country's economic welfare. This happens if, and only if, imports are capital intensive in the case of capital movement, or if, and only if, technological progress is capital saving in either industry, or Hicksian neutral in the import-competing industry in the case of technology movement. When the export industry uses foreign technology of the Hicksian-neutral

progress type, the output of the import-competing industry does not change at all (though that of the export industry increases and will be repatriated) at constant terms of trade, so that the home country's economic welfare is not affected.

In the case of capital movement, all these results are applicable to the non-fixed-production coefficient case so long as the home country produces both F and M. However, as already stated, this is not true in the case of technology movement. For instance, when the export industry uses foreign technology of the Hicksian-neutral progress type, the home country's economic welfare improves in the non-fixed-production coefficient case, because the domestic supply of the import-competing good falls at a constant domestic commodity price ratio, and imports increase.

When the country specializes completely in a single industry, a small amount of capital inflow or technology inflow will not increase the home country's economic welfare in the fixed-production coefficient case, because the country cannot shift resources out of the import-competing goods into exports (notice that there is no output of the import-competing good before capital or technology moves) and there is no substitution of capital for labor. However, when the possibility of substitution of capital for labor is allowed, gains can increase after complete specialization in the capital-intensive good sets in because then the return to foreign capital or foreign technology of the capital-saving type is depressed and intramarginal units are paid less than their contribution to output. But the same kind of capital or technology inflow will reduce the home country's economic welfare when the country specializes completely in the labor-intensive good, because the country begins to produce the import-competing good at the import price plus the tariff, and imports decrease.

REFERENCES

H. G. Johnson, "The Possibility of Income Losses from Increased Efficiency or Factor Accumulation in the Presence of Tariffs," *Econ. J.*, Mar. 1967, *77*, 151–54.

L. Metzler, "Tariffs, the Terms of Trade and Distribution of National Income," *J. Polit. Econ.*, Feb. 1949, *57*, 1–29.

N. Minabe, "The Heckscher-Ohlin Theorem, the Leontief Paradox, and Patterns of Economic Growth," *Amer. Econ. Rev.*, Dec. 1966, *56*, 1193–1211.

———, "The Linear-Programming Structure of the Pure Theory of International Trade and Intermediate Goods," unpublished, 1974.

R. A. Mundell, "International Trade and Factor Mobility," *Amer. Econ. Rev.*, June 1957, *47*, 321–35.

T. Rybczynski, "Factor Endowment and Relative Commodity Prices," *Economica*, Nov. 1955, *22*, 336–41.

W. F. Stolper and P. A. Samuelson, "Protection and Real Wages," *Rev. Econ. Stud.*, Nov. 1941, *9*, 58–73.

H. Uzawa, "Shihonjiyuka to Kokuminkeizai (Capital Liberalization and the National Economy)," *Japan Economist*, Dec. 1969, *23*, 106–22.

ERRATA

The Foundations of Money Illusion in a Neoclassical Micro-Monetary Model

By RICHARD DUSANSKY AND PETER J. KALMAN

Please note the following revisions in the March 1974 issue of this *Review:*

Page 115:

equation (1) should read: $u = u(x_1, \ldots, x_n, P)$
column 2, line 6 should read: "nominal" instead of "real."
column 2, line 18 should read: $u(x_1, \ldots, x_n, P)$
column 2, line 20 should read: "second" instead of "first."
column 2, line 23: delete "more."

Page 121:

footnote 8, add: and *Erratum*, same journal, *6*, p. 107 (1973).

Page 122:

equation (26) should read:

$$\sum_{j=1}^{n-1} \frac{\partial m}{\partial p_j} p_j + \frac{\partial m}{\partial y} y + \frac{\partial m}{\partial L} L = m$$

equation (28) should read:

$$\sum_{j=1}^{n-1} \frac{\partial m}{\partial p_j}\left(\frac{p_j}{m}\right) + \frac{\partial m}{\partial y}\left(\frac{y}{m}\right) + \frac{\partial m}{\partial L}\left(\frac{L}{m}\right) = 1$$

column 2, last sentence should read: " . . . must appropriately sum to zero and unity, respectively, is empirically refutable."

IN MEMORIAM

HAROLD HOTELLING

1895–1973

Harold Hotelling, creative thinker in both mathematical statistics and economics, was born in Fulda, Minnesota, 29 September 1895, and died in Chapel Hill, North Carolina, on 26 December 1973. His influence on the development of economic theory was deep, though it occupied a relatively small part of a highly productive scientific life devoted primarily to mathematical statistics; only ten of some eighty-seven published papers were devoted to economics, but of these, six are landmarks which continue to this day to lead to further developments. His major research on mathematical statistics had, further, a generally stimulating effect on the use of statistical methods in different specific fields of application, including econometrics; but more strikingly his papers on principal components and canonical correlation were closely associated with the development of the estimation of simultaneous equations.

His early interests were in journalism; he received his B.A. in that field from the University of Washington in 1919. Later in classes, he would illustrate the use of dummy variables in regression analysis by a study (apparently never published) of the effect of the opinions of different Seattle newspapers on the outcome of elections and referenda. The mathematician and biographer of mathematicians, Eric T. Bell, discerned talent in Hotelling and encouraged him to switch his field. He received an M.A. in mathematics at Washington in 1921 and a Ph.D. in the same field from Princeton in 1924; he worked under the topologist, Oswald Veblen (Thorstein Veblen's nephew); and two of his early papers dealt with manifolds of states of motion.

The year of completing his Ph.D., he joined the staff of the Food Research Institute at Stanford University with the title of Junior Associate. In 1925, he published his first three papers: one on manifolds, one on a derivation of the F-distribution, and one on the theory of depreciation. Here, apparently for the first time, he stated the now generally accepted definition of depreciation as the decrease in the discounted value of future returns. This paper was a turning point both in capital theory proper and in the reorientation of accounting towards more economically meaningful magnitudes.

In subsequent years at Stanford he became research associate of the Food Research Institute and associate professor of mathematics, teaching courses in mathematical statistics and probability (including an examination of Keynes's *Treatise on Probability*) along with others in differential geometry and topology. In 1927, he showed that trend projections of population were statistically inappropriate and introduced the estimation of differential equations subject to error; he returned to the statistical interpretation of trends in 1929 in a notable joint paper with Holbrook Working, largely under the inspiration of the needs of economic analysis.

The same year he published the famous paper on stability in competition, in which he introduced the notions of locational equilibrium in duopoly. This paper is still anthologized and familiar to every theoretical economist. As part of the paper, he noted that the model could be given a political interpretation, that competing parties will tend to have very similar programs. Although it took a long time for subsequent models to arise, these few pages have become the source for a large and fruitful literature.

In 1931, his paper on the economics of exhaustible resources appeared, an application of the calculus of variations to the problem of allocation of a fixed stock over time. All of the current literature, inspired by the growing sense of scarcity (natural and arti-

ficial), is essentially based on Hotelling's paper. Interestingly enough, according to his later accounts, the *Economic Journal* rejected the paper because its mathematics were too difficult (although it had published Ramsey's paper earlier); it was finally published in the *Journal of Political Economy*.

The same year, he was appointed professor of economics at Columbia University, where he was to remain until 1946. There he began the organization of a systematic curriculum in theoretical statistics which eventually attained the dignity of a separate listing in the catalogue, though not the desired end of a department or degree-granting entity. Toward the end of the 1930's, he attracted a legendary set of students who represented the bulk of the next generation of theoretical statisticians. His care and encouragement of his students were extraordinary: the encouragement of the self-doubtful, the quick recognition of talent, the tactfully made research suggestion at crucial moments created a rare human and scholarly community. He was as proud of his students as he was modest about his own work.

He also gave a course in mathematical economics. The general environment was not too fortunate. The predominant interests of the Columbia department of economics were actively antitheoretical, to the point where *no* systematic course in neoclassical price theory was even offered, let alone prescribed for the general student. Nevertheless, several current leaders in economic theory had the benefit of his teaching. But his influence was spread more through his papers, particularly those on the full development of the second-order implications for optimization by firms and households (contemporaneous with Hicks' and Allen's papers) and above all by his classic presidential address (1938) before the Econometric Society on welfare economics. Here we have the first clear understanding of the basic propositions (Hotelling, as always, was meticulous in acknowledging earlier work back to Dupuit), as well as the introduction of extensions from the two-dimensional plane of the typical graphical presentation to the calculation of benefits with many related commodities. Here also we have the clearest expression in print of Hotelling's strong social interests which motivated his technical economics. His position was undogmatic but in general it was one of market socialism. He had no respect for acceptance of the status quo as such, and the legitimacy of altering property rights to benefit the deprived was axiomatic with him; but at the same time he was keenly aware of the limitations on resources and the importance in any human society of the avoidance of waste.

In 1946, he finally had the long-desired opportunity of creating a department of mathematical statistics at the University of North Carolina, where he remained until retirement. He continued his active interest in economics there.

Space forbids more than the brief mention of his important work in the foundation of two learned societies, the Econometric Society and the Institute of Mathematical Statistics, both of which he served as president at a formative stage. He received many formal honors during his lifetime, including honorary degrees from Chicago and Rochester; he was the first Distinguished Fellow of the American Economic Association when that honor was created. To him, the admiration and love of his former students, repeatedly expressed, was probably even a greater honor.

He is survived by Mrs. Susanna Hotelling (whose gracious and engaging role as hostess of the Hotellings' monthly Sundays for students is appreciated by generations of his students) and seven children.

NOTES

Starting January 1, 1974, all manuscripts submitted to this *Review* have been sent out to referees without the author's name or affiliation. Original submissions should have a separate cover page for the title of the paper. For purposes of refereeing, authors are requested to avoid self-identification in the main text of their papers.

Annual Meeting Placement Service

The Placement Service at the 1974 annual meetings of the Allied Social Science Associations in San Francisco will begin operation on December 27, the day before sessions begin. Applicants and employers will be able to attend more sessions with a day set aside entirely for labor market transactions. The Placement Service will continue during the other days of the annual meetings, but hopefully at a less frantic pace than in previous years.

The 1975 annual meetings of the American Economic Association will be held in Dallas, October 3–5, not October 2–4 as previously announced. The Placement Service will begin operations on October 2.

The program chairman of the American Economic Association has decided to have several sessions of contributed papers at the October 1975 meetings of the Association in Dallas. It is expected that some of these sessions will be held jointly with the Econometric Society. Those interested in submitting papers for consideration are invited to send abstracts *not later than February 1* to Franco Modigliani, Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139. Authors should indicate whether they would like their paper to be considered for a joint session with the Econometric Society. The abstract should not exceed two typewritten pages. A group of papers will be selected from the abstracts submitted.

The Executive Committee of the American Economic Association at its March 1974 meeting voted to increase the dues of regular members from $21.00 to $23.00, and of other members proportionately effective January 1, 1975. This action was taken under the provision of the bylaws permitting the Executive Committee to increase dues proportionately to increases in relevant costs and prices.

At its meeting on December 27, 1973, the Executive Committee of the American Economic Association approved a proposal for a Visiting Economics Scholars Program by Edward J. Powers of Northern Michigan University. The program will be modeled on the Visiting Scientists Program formerly funded by the National Science Foundation and will begin with the academic year 1974–75. The goal is to provide visiting scholars to the more geographically isolated and less affluent colleges and universities across the country. For more information, contact C. Elton Hinshaw, Assistant Secretary-Treasurer, American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212.

The Joint Committee on Soviet Studies and the Joint Committee on Eastern Europe of the American Council of Learned Societies and the Social Science Research Council wish to draw attention to Grants Programs in Soviet Studies and East European Studies. The programs include Grants for Post-Doctoral Research, Grants for Study of East European Languages, Grants in Support of Conferences, and Travel Grants to International Conferences Abroad. For details of eligibility and information which *must* be supplied in requesting application forms, request an announcement brochure from Office of Fellowships and Grants, American Council of Learned Societies, 345 E. 46th Street, New York, New York 10017.

Old Dominion University and Norfolk State College invite participation in the Virginia Conference on Urban Studies to be held on February 13 and 14, 1975 at Old Dominion University. The Conference is designed to bring together urban specialists from all the social sciences in a series of meetings focussing on "The Urban South." For further information, write Professor Harold Wilson, Department of History, Old Dominion University, Norfolk, Virginia, or Professor Charles Simmons, Department of History, Norfolk State College, Norfolk Virginia.

The seventh annual meeting of The International Society for the History of Behavioral and Social Sciences (CHEIRON) will be held at Carleton University on June 5–8, 1975. Papers will be considered which deal with aspects of the history of any of the behavioral and social sciences, with related historical or social science methodology, or with the philosophy of history as applied to the study of the history of the behavioral and social sciences. The emphasis of the meeting will be interdisciplinary. Address submitted papers (deadline Jan. 1, 1975) or requests for additional information to Dr. M. E. Marshall, Department of Psychology, Carleton University, Ottawa, Ontario, Canada K1S. For information concerning membership, write Dr. Elizabeth S. Goodman, 115 West Royal Drive, DeKalb, Illinois 60115.

On Thursday, May 15, 1975, the Economics Department of the City College of New York will sponsor an

all day conference on "The Economics of Scarce Resources." The object of the conference is to stimulate research that is relevant to current public issues and policies. Persons interested in reading papers at the conference are invited to submit them to Professor Robert D. Leiter or Professor Stanley Friedlander, Department of Economics, City College of the City University of New York, Convent Avenue and West 138 Street, New York, N.Y. 10031. Papers delivered at the conference are scheduled for publication in the third annual conference volume of the department.

The Council for European Studies, a consortium of U.S. universities, will sponsor a pre-dissertation fellowship program with an application deadline of February 1, 1975, and a workshop program in 1975. Inquiries are invited from all social sciences and related disciplines and are particularly encouraged from those areas of study in which relatively less attention has been given to Western Europe, such as economics, sociology, anthropology, and social psychology. A limited amount of support resources is available. For more information and application forms, graduate students and younger postdoctoral scholars should contact the Council for European Studies, 156 Mervis Hall, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

History of Economics Society

The purpose of the History of Economics Society is to promote interest and inquiry into the history of economics and related parts of intellectual history; facilitate communication and discourse among scholars working in the field; and acquaint members of the profession with the scientific, literary, and philosophical tradition of economics. Information about membership in the Society and about a reduced-price subscription to the quarterly journal *History of Political Economy* can be obtained from James L. Cochrane, Secretary-Treasurer, History of Economics Society, University of South Carolina, Columbia, South Carolina 29208. The Society will hold its next annual conference at Baker Library, Harvard University on May 21–24, 1975. Persons interested in participating at the conference should contact Dr. Joseph J. Spengler, President-Elect, History of Economics Society, Department of Economics, Duke University, Durham, North Carolina 27706.

Omicron Delta Epsilon, the International Honor Society in Economics, invites the submission of entries for the Irving Fisher Graduate Monograph and Frank W. Taussig Award Article Competitions. The Fisher Award consists of $1,000 and publication as a book by the Princeton University Press, subject to approval of its editorial board. In addition, the winner will be invited to submit a paper based on the winning entry to the *American Economic Review*. The recommendations of the Final Selection Board of the Competition will be considered by the *Review* in the refereeing process. All finalists will be invited to submit a paper for publication in *The American Economist*. The Taussig Award consists of $100 and publication in *The American Economist*. Entries for the Fisher Award should be submitted to departmental selection committees by January 1, 1975 and entries for the Taussig Award by May 15, 1975. They will be judged by the International Editorial Board and finalists by the Final Selection Board, consisting of Professors William J. Baumol, Frank H. Hahn, Robert M. Solow, Arnold Zellner, and Egon Neuberger (editor). For more information, write Professor Edward Ames, Acting Editor, Economic Research Bureau, State University of New York, Stony Brook, New York 11794.

The American Iron and Steel Institute announces the availability of several fellowships for advanced graduate study in economics, management, and related disciplines. Your attention is called to a full page advertisement in the ad section of this issue.

Papers are solicited for the seventh Amos Tuck School Seminar on Problems of Regulation and Public Utilities, sponsored by the American Telephone and Telegraph Company, to be held in early August of 1975 at Dartmouth College. The seminar will be attended by thirty conferees chosen from academic institutions throughout the country. It consists of seven or eight sessions over a four-day period. The sessions provide a forum for the presentation and discussion of new ideas and innovative work in analyzing regulated firms, regulatory processes, and problems of public welfare.

We invite those with an interest in this area to suggest names of possible speakers or to apply themselves. The aim of the seminar is to introduce the research both of young scholars beginning their careers and of scholars who have only recently turned their attention to public utility problems. Therefore, suggestions of speakers must be limited to scholars in these categories. Those chosen to present papers will receive support for research and writing, in addition to travel and living expenses at the seminar. Please send suggestions to the directors of the seminar before February 7, 1975. Include two copies of each of the following: biographical information, samples of written work, and a proposal to be developed for presentation. This is not a request for conferee applications. Application forms and further information will be mailed to institutions in March, 1975. The directors of the seminar are Professors Duncan M. Holthausen, Jr., and Dennis E. Logue. Please direct all suggestions and correspondence to Directors, Seminar on Problems of Regulation and Public Utilities, The Amos Tuck School of Business Administration, Dartmouth College, Hanover, New Hampshire 03755.

Deaths

Persia Campbell, professor emeritus, Queens College, City University of New York, Mar. 2, 1974.

Addison T. Cutler, professor of economics emeritus, University of Cincinnati, May 24, 1974.

Helen C. Farnsworth, professor emeritus and former associate director, Food Research Institute, Stanford University, Feb. 23, 1974.

Clyde O. Fisher, professor emeritus, Wesleyan University, Middletown, Connecticut, June 15, 1974.

Retirements

Arthur E. Burns, dean, George Washington University Graduate School of Arts and Sciences, June 1974.

Joseph D. Coppock, professor of economics, Pennsylvania State University: visiting professor, American University of Beirut, Lebanon, 1974–76.

Walter P. Egle, professor of economics emeritus, University of Cincinnati, Aug. 31, 1974.

Marion H. Gillim, Barnard College: professor of economics, Eastern Kentucky University, Aug. 1974.

John A. Hogan, professor of economics, Whittemore School of Business and Economics, University of New Hampshire, July 1, 1974.

Ernest M. Stallings, department of economics, North Carolina State University, June 1974.

G. Carl Wiegand, professor of economics, Southern Illinois University, Aug. 31, 1974.

William Withers, professor of economics, Queens College, City University of New York, Feb. 1, 1974.

Visiting Foreign Scholars

Paul Newbold, University of Nottingham, England: visiting professor of economics, University of California, San Diego, 1974–75.

Rodney D. Peterson, Colorado State University: visiting professor, department of economics and commerce, Simon Fraser University, Sept. 1974.

J. Ernest Tanner, Tulane University: visiting professor, department of economics and commerce, Simon Fraser University.

Promotions

Richard D. Bartel: special assistant, foreign department, Federal Reserve Bank of New York, Apr. 4, 1974.

Stanley M. Besen: professor of economics, Rice University, July 1974.

Roger R. Betancourt: associate professor of economics, University of Maryland, Aug. 1974.

Richard A. Bilas: professor of business and economics, California State College, Bakersfield, July 1974.

Frank J. Bonello: associate professor, University of Notre Dame, 1974–75.

W. Robert Brazelton: professor, department of economics, University of Missouri-Kansas City, Sept. 1974.

Malcolm C. Brown: associate professor, department of economics, University of Calgary, July 1, 1974.

J. Bruce Bullock: associate professor of economics, North Carolina State University, July 1, 1974.

Gerald A. Carlson: associate professor of economics, North Carolina State University, July 1, 1974.

John Conlisk: professor of economics, University of California, San Diego, July 1, 1974.

D. A. Dawson: associate professor of economics, McMaster University.

Edward J. Deak: associate professor of economics, Fairfield University, Sept. 1974.

Robert F. Dernberger: professor, department of economics, University of Michigan, July 1, 1974.

James E. Easley, Jr.: extension assistant professor of economics, North Carolina State University, July 1, 1974.

Kenneth G. Elzinga: professor of economics, University of Virginia, Sept. 1974.

Robert F. Engle: associate professor, department of economics, Massachusetts Institute of Technology.

Edward W. Erickson: professor of economics, North Carolina State University, July 1, 1974.

Terry Farrar: associate professor of economics, Pennsylvania State University, July 1, 1974.

Maryann Fogarty: instructor, department of economics, Barnard College, Sept. 1974.

Arthur M. Ford: associate professor of economics, Southern Illinois University, Aug. 16, 1974.

Raymond S. Franklin: professor of economics, Queens College, City University of New York.

Christopher Garbacz: associate professor of economics, University of Missouri-Rolla, Aug. 1974.

Henry C. Gilliam, Jr.: associate professor of economics, North Carolina State University, July 1, 1974.

R. Jeffery Green: associate professor of economics, Indiana University, July 1, 1974.

Margaret L. Greene: assistant vice president, foreign function, Federal Reserve Bank of New York, Feb. 22, 1974.

Arthur Guthrie: associate professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1974.

Robert E. Hall: professor of economics, Massachusetts Institute of Technology.

Daryl A. Hellman: associate professor, department of economics, Northeastern University, fall 1974.

Timothy Hogan: associate professor of economics, Arizona State University.

James O. Horrigan: professor of business administration, Whittemore School of Business and Economics, University of New Hampshire, Sept. 1, 1974.

J. A. Johnson: professor of economics, McMaster University.

Charles P. Jones: associate professor of economics, North Carolina State University, July 1, 1974.

James L. Kenkel: associate professor, University of Pittsburgh, Sept. 1, 1974.

M. L. Kliman: associate professor of economics, McMaster University.

Roger Kubarych: foreign exchange officer, foreign department, Federal Reserve Bank of New York, July 1, 1974.

A. A. Kubursi: associate professor of economics, McMaster University.

Patricia H. Kuwayama: chief, foreign research division, Federal Reserve Bank of New York, July 25, 1974.

James C. Loughlin: professor of economics, Central Connecticut State College.

Dennis A. MacDonneil: head, international management development and training, Bank of America, San Francisco, June 18, 1974.

Rodney H. Mabry: assistant professor of economics, Clemson University, May 16, 1974.

Don H. Mann: assistant professor, department of

economics and commerce, Simon Fraser University, Sept. 1, 1974.

G. Frank Mathewson: associate professor of economics, University of Toronto, July 1, 1974.

Wolfgang Mayer: associate professor of economics, University of Cincinnati, Sept. 1, 1974.

James P. Moody: assistant professor, department of economics, University of Wisconsin-Milwaukee.

George Neumann: associate professor of economics, Pennsylvania State University, July 1, 1974.

Christos C. Paraskevopoulos: associate professor, department of economics, York University, July 1, 1974.

Scott E. Pardee: vice president, foreign function, Federal Reserve Bank of New York, Feb. 22, 1974.

Lawrence T. Pinfield: associate professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1974.

Julius C. Poindexter, Jr.: associate professor of economics, North Carolina State University, July 1, 1974.

Robert C. Puth: associate professor of economics, Whittemore School of Business and Economics, University of New Hampshire, Sept. 1, 1974.

Edward J. Ray: associate professor of economics, Ohio State University.

R. Robert Russell: professor of economics, University of California, San Diego, July 1, 1974.

Leonard G. Sahling: chief, business conditions division, Federal Reserve Bank of New York, May 16, 1974.

Pawan K. Sawhney: assistant professor, department of economics, Northeastern University, fall 1974.

Richard L. Schmalensee: associate professor of economics, University of California, San Diego, July 1, 1974.

Ronald A. Schrimper: professor of economics, North Carolina State University, July 1, 1974.

Richard Schwindt: assistant professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1974.

Sureh P. Sethi: associate professor, University of Toronto, July 1, 1974.

Inderjit Singh: associate professor of economics, Ohio State University.

Roger Skurski: associate professor, University of Notre Dame, 1974–75.

Gordon W. Smith: associate professor of economics, Rice University, July 1974.

John Soladay: assistant professor of economics, Pennsylvania State University, July 1, 1974.

Zane A. Spindler: associate professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1974.

Nicolas Spulber: professor of economics, Indiana University, July 1, 1974.

Ernst W. Stromsdorfer: professor of economics, Indiana University, July 1, 1974.

Michael Szenberg: professor of economics, Long Island University, September 1974.

Richard E. Towey: professor, department of economics, Oregon State University, Sept. 16, 1974.

Martin L. Weitzman: professor of economics, Massachusetts Institute of Technology.

Robert C. Wells: extension professor of economics, North Carolina State University, July 1, 1974.

J. R. Williams: professor of economics, McMaster University.

Gavin Wright: professor, department of economics, University of Michigan, July 1, 1974.

Holley H. Ubrich: associate professor of economics, Clemson University, May 16, 1974.

Oleg Zinam: professor of economics, University of Cincinnati, Sept. 1, 1974.

Administrative Appointments

Douglas R. Bohi: chairman, department of economics, Southern Illinois University, Aug. 16, 1974.

Joseph L. Craycraft: director of Graduate Studies, department of economics, University of Cincinnati, Sept. 1, 1974.

John S. Curtiss: director of undergraduate studies, department of economics, University of Cincinnati, Sept. 1, 1974.

Ana N. Eapen: chairperson, department of economics and business, William Paterson College, July 1, 1974.

Terry A. Ferrar: director, Center for the Study of Environmental Policy, Pennsylvania State University, July 1, 1974.

Peter J. George: associate dean of graduate studies, McMaster University, 1974–77.

William F. Hellmuth, McMaster University: chairman, department of economics, Virginia Commonwealth University, Dec. 1973.

E. Edward Herman: head, department of economics, University of Cincinnati, Sept. 1, 1974.

Hal Hoverland: dean, School of Administration, California State College, San Bernardino, July 1, 1974.

William H. Leahy: acting chairman, department of economics, University of Notre Dame, Sept. 1, 1974.

R. Craig McIvor: acting dean of social sciences, McMaster University, 1974–75.

Deborah D. Milenkovitch: chairman, department of economics, Barnard College, July 1, 1974.

Jerry Miner: chairman, department of economics, Syracuse University, June 1, 1974.

Clair E. Morris: chairman, department of economics, U.S. Naval Academy, July 1974.

John D. Pilgrim, Bradley University: chairman and associate professor, department of economics, Ursinus College, Sept. 1974.

W. Phillip Saunders, Jr.: associate dean, College of Arts and Science, Indiana University, July 1, 1974.

Don A. Seastone: head, department of economics, University of Calgary, July 1, 1974.

Harold T. Shapiro: chairman, department of economics, University of Michigan, July 1, 1974.

Lester P. Silverman: associate executive director, Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences.

Peter N. Vukasin, State University of New York at Binghamton: vice president for academic affairs, State University of New York at New Paltz, Sept. 1, 1974.

Armand J. Zottola: chairman, department of economics, Central Connecticut State College.

Appointments

Lawrence W. Abrams, Washington University: acting assistant professor of economics, Board of Studies in Economics, University of California at Santa Cruz, July 1974.

C. Michael Aho: instructor, department of economics, Northeastern University, fall 1974.

Richard J. Arnott, Yale University: visiting instructor in economics, Wesleyan University, Sept. 1, 1974.

Robert Avanian: visiting assistant professor of economics, Arizona State University.

Martin J. Bailey: professor of economics, University of Maryland, Aug. 1974.

Ann Bartel: research associate, National Bureau of Economic Research, July 1, 1974.

Carson Bays: assistant professor of economics, University of Illinois at Chicago Circle, Sept. 1974.

Gregory Boussios, Fordham University: assistant professor, William Paterson College, Sept. 1974.

William Boyes: assistant professor of economics, Arizona State University.

F. Michael Bradfield, Dalhousie University: visiting assistant professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1974.

Dale G. Broderick: associate professor of business administration, Whittemore School of Business and Economics, University of New Hampshire, Sept. 1, 1974.

Oscar T. Brookins, University of Ghana, Africa: assistant professor of economics, University of Notre Dame, 1974–75.

J. B. Burbidge: assistant professor, department of economics, McMaster University.

John M. Burt, Jr.: assistant professor of business administration, Whittemore School of Business and Economics, University of New Hampshire, Sept. 1, 1974.

John N. Carvellas, Syracuse University Research Corporation: instructor of economics, St. Michael's College, Sept. 1974.

Paul P. Christensen, University of Wisconsin, Madison: lecturer in economics, Board of Studies in Economics, University of California at Santa Cruz.

Sandra S. Christensen, University of Maryland: assistant professor, department of economics and commerce, Simon Fraser University, Sept. 1, 1974.

Hyun Sik Chung: assistant professor, department of economics, Northeastern University, fall 1974.

Lloyd Coen, State University of New York at Binghamton: assistant professor, department of economics, Laurentian University, Sept. 1974.

Charles Conrod, Simon Fraser University: research associate, Worker's Compensation Board of British Columbia, Apr. 1974.

George M. Constantinides, University of Indiana: assistant professor, Carnegie-Mellon Uniersity, Sept. 1, 1974.

Robert W. Crandall: Center for Policy Alternatives, Massachusetts Institute of Technology, June 30, 1974.

Omar Davies: assistant professor, Food Research Institute, Stanford University, Sept. 1974.

Gopal Dorai, Wheaton College: associate professor, department of economics, William Paterson College, Sept. 1974.

Michael Edelstein, Columbia University: assistant professor, Queens College, City University of New York, 1974–75.

Barry Edmonston: acting assistant professor and research demographer, Food Research Institute, Stanford University, Dec. 1973.

Hugh N. Emerson, University of New Haven: resident lecturer, University of Maryland, University College, Atlantic Division (the Azores, Bermuda, Iceland, Labrador), June 1, 1974.

Dennis Epple, Princeton University: assistant professor, Carnegie-Mellon University, Sept. 1, 1974.

J. Eric Fredland, George Washington University: assistant professor, U.S. Naval Academy, Aug. 1974.

Ann F. Friedlaender: professor of economics, Massachusetts Institute of Technology, Sept. 1, 1974.

Shelby Gerking: visiting assistant professor of economics, Arizona State University.

Sharif Ghalib: assistant professor of economics, Pennsylvania State University, Sept. 1, 1974.

Arthur Gibb, Lawrence University: assistant professor, U.S. Naval Academy, Aug. 1974.

Harvey N. Gram, University of Waterloo: assistant professor, Queens College, City University of New York, 1974–75.

Elizabeth F. Gustafson: assistant professor of economics, University of Cincinnati, Sept. 1, 1974.

James F. Halstead, Williams College: lecturer in economics, Board of Studies in Economics, University of California at Santa Cruz, July 1974.

Richard E. Hamilton, OECD, Paris: associate professor, department of economics, University of Calgary, Sept. 1, 1974.

Milton Harris, University of Chicago: assistant professor, Carnegie-Mellon University, Sept. 1, 1974.

Joseph A. Hasson: professor of economics, Howard University.

Walter P. Heller, University of Pennsylvania: associate professor of economics, University of California at San Diego, July 1, 1974.

Chris J. Henry: instructor, department of economics, Northeastern University, fall 1974.

Sylvia A. Hewlett: assistant professor of economics, Barnard College, Sept. 1, 1974.

Alexander Holmes, State University of New York at Binghamton: assistant professor, department of economics, University of Oklahoma, Sept. 1, 1974.

Joseph Hu: assistant professor of economics, University of Maine, Sept. 1, 1974.

Hillard G. Huntington, State University of New York at Binghamton: staff economist, Federal Energy Administration, Washington, Sept. 1, 1974.

M. Edwin Ireland, Texas Tech University: assistant professor of economics, Clemson University, Aug. 15, 1974.

William H. Janeway: adjunct assistant professor of economics, Barnard College, Sept. 1, 1974.

William R. Johnson, Wesleyan University: assistant professor of economics, University of Virginia, Sept. 1, 1974.

T. David Johnston: assistant professor of economics, Indiana University Southeast, Aug. 1974.

Maryann O. Keating: assistant professor of economics, Radford College, Sept. 1974.

Harry Kelejian, New York University: professor of economics, University of Maryland, Aug. 1974.

J. Douglas Klein, University of Wisconsin, Madison: assistant professor of economics, Syracuse University, Sept. 1, 1974.

Bruce L. Kutnik: instructor, department of economics, Northeastern University, fall 1974.

Edward P. Lazear: research associate, National Bureau of Economic Research, June 1, 1974.

Edward Lesnick, Jr., Valparaiso University: assistant professor, department of economics, William Paterson College, Sept. 1974.

Robert A. Lewis, Ohio University: assistant professor of urban and regional planning, Virginia Commonwealth University, Sept. 1, 1974.

K. Lynch: lecturer, department of economics, McMaster University, 1974–75.

Vern D. Lyon, University of Kansas: visiting assistant professor, Southern Illinois University, Aug. 16, 1974.

Robert E. McCormick: instructor of economics, Clemson University, Aug. 15, 1974.

Christopher J. McCurdy: economist, Money and Finance Division, Federal Reserve Bank of New York, Apr. 16, 1974.

Michael T. Maloney, Louisiana State University: instructor of economics, Clemson University, Aug. 15, 1974.

Judith K. Mann, University of Hawaii: assistant professor of economics, University of California at San Diego, July 1, 1974.

Patricio Meller: research fellow, National Bureau of Economic Research, Jan. 1, 1975.

Paul L. Menchik: lecturer, department of economics, Rutgers—The State University, Sept. 1974.

I. Dennis Menezes: assistant professor of business administration, Indiana University Southeast, Aug. 1974.

Daniel C. Messerschmidt, Iowa State University: visiting assistant professor, Southern Illinois University, Aug. 16, 1974.

Tapan Mitra: assistant professor of economics, University of Illinois at Chicago Circle, Sept. 1974.

John F. Moeller: assistant professor of economics, University of Maine, Sept. 1, 1974.

John E. Morton, Board of Governors, Federal Reserve System: assistant professor of economics, University of Maryland, Aug. 1974.

Bela Mukhoti, University of Northern Iowa: associate professor, Glassboro State College, 1974–75.

Willis J. Nordlund: assistant professor, department of economics, Oregon State University, Sept. 16, 1974.

Maureen O'Brien, Oklahoma State University: instructor of economics, Clemson University, Aug. 15, 1974.

James F. O'Connor: assistant professor of economics, Pennsylvania State University, Sept. 1, 1974.

Thomas J. Parliment, State University of New York at Binghamton: assistant professor, department of economics, University of Wisconsin-Parkside, Sept. 1, 1974.

Sam Peltzman: research associate, National Bureau of Economic Research, Sept. 1, 1974.

Joseph Persky: associate professor of economics, University of Illinois at Chicago Circle, Sept. 1974.

John G. Pomery, University of Rochester: assistant professor of economics, Rice University, July 1, 1974.

Robert Posatko: assistant professor of economics, Pennsylvania State University, Sept. 1, 1974.

Burt E. Powell, Duke University: assistant professor, department of economics, Virginia Commonwealth University, Aug. 1974.

Artur Raviv, Northwestern University: assistant professor, Carnegie-Mellon University, Sept. 1, 1974.

Carl A. Riskin, Columbia University: assistant professor, Queens College, City University of New York, 1974–75.

John A. Robson: assistant professor, department of economics, Northeastern University, fall 1974.

Elizabeth A. Roistacher, University of Michigan: assistant professor, Queens College, City University of New York, 1974–75.

Terry L. Rush, University of Oklahoma: assistant professor, department of economics, Virginia Commonwealth University, Aug. 1974.

Mahmoud Sakbani: economist, Balance of Payments Division, Federal Reserve Bank of New York, Mar. 28, 1974.

James E. Sawyer: assistant professor, department of economics, Oregon State University, Sept. 16, 1974.

Dewey Q. Seeto: assistant professor, department of economics, Rutgers—The State University, Sept. 1974.

Eugene Seskin: research associate, National Bureau of Economic Research, May 1, 1974.

Russell F. Settle, University of Wisconsin-Madison: assistant professor, department of economics, University of Delaware, Sept. 1974.

Bradley R. Schiller, University of Maryland: lecturer in economics, Board of Studies in Economics, University of California, Santa Cruz, July 1974.

Mary E. Scovill: assistant professor of economics, University of Notre Dame, 1974–75.

Sue O. Shaw: visiting assistant professor of economics, Clemson University, Aug. 15, 1974.

Ralph Shlomowitz: lecturer of economics, University of Illinois at Chicago Circle, Sept. 1974.

Barry Shore: associate professor of business administration, Whittemore School of Business and Economics, University of New Hampshire, Sept. 1, 1974.

Andrew Shotter, New York University: assistant professor of economics, Syracuse University, Sept. 1, 1974.

Ross M. Starr: associate professor, department of economics, University of California, Davis, Jan. 1, 1975.

Robert R. Sterling, University of Kansas: professor of accounting, Rice University, Feb. 1974.

John Stevens: assistant professor of economics, Pennsylvania State University, Sept. 1, 1974.

Mo-Yin Tam: assistant professor of economics, University of Illinois at Chicago Circle, Sept. 1974.

Lance J. Taylor: professor of nutritional economics, Massachusetts Institute of Technology, Sept. 1, 1974.

Allen R. Thompson: assistant professor of economics, Whittemore School of Business and Economics, University of New Hampshire, Sept. 1, 1974.

Ronald L. Tracy, Michigan State University: assistant professor, Southern Illinois University, Aug. 16, 1974.

Hiroki Tsurumi: associate professor, department of economics, Rutgers—The State University, Sept. 1974.

Jerald H. Udinsky, University of California, Berkeley. assistant professor of economics, Board of Studies in Economics, University of California at Santa Cruz, July 1974.

E. Lane Vanderslice, Rutgers—The State University: assistant professor of economics, University of Notre Dame, 1974–75.

Frank M. Werner: lecturer, department of economics, Rutgers—The State University, Sept. 1974.

James N. Wetzel, University of North Carolina, Greensboro: assistant professor, department of economics, Virginia Commonwealth University, Aug. 1974.

John R. Woodbury, Washington University: lecturer, Southern Illinois University, Aug. 16, 1974.

Hassan Zavareei, Trenton State College: assistant professor of economics, West Virginia Institute of Technology, Aug. 15, 1974.

Allan Zelenitz, Princeton University: assistant professor, University of Pittsburgh, Sept. 1, 1974.

Leaves for Special Appointments

Charles Blackorby, Southern Illinois University: visiting associate professor, University of British Columbia, Aug. 16, 1974–May 15, 1975.

John P. Bonin, Wesleyan University: visiting assistant professor of economics, University of California, San Diego, Sept. 1, 1974.

Stephen A. Buser, Southern Illinois University: visiting research economist, Federal Deposit Insurance Corporation, Aug. 16, 1974–May 15, 1975.

Garrey E. Carruthers, New Mexico State University: White House Fellow, Washington, 1974–75.

Thomas G. Cowing, State University of New York at Binghamton: Brookings Economic Policy Fellow, Environmental Protection Agency, Washington, Sept. 1, 1974.

Joseph S. Desalvo, University of Wisconsin-Milwaukee: visiting scholar, faculty of economic science, Catholic University of Mons, Belgium, Aug. 1, 1974.

Peter G. Franck, Syracuse University: visiting professor, National University of Zaire, Kinshasa, Oct. 1, 1974.

Nancy Gordon, Carnegie-Mellon University: Brookings Fellow, Department of Labor, June 1, 1974.

Kenneth P. Jameson, University of Notre Dame: Fulbright Lecturer, University of San Agustin, Areguipa, Peru, 1974–75.

Bruce F. Johnston, Stanford University: senior research fellow, Institute for Development Studies, University of Nairobi, Kenya, Apr. 1974–Sept. 1975.

Jerry Ladman, Arizona State University: Fulbright Lecturer, University of Guayquil, Educador.

Robert G. Layer, Southern Illinois University: Fulbright-Hays Lecturer, Faculty of Public and Business Administration, University of Tehran, Iran, Sept. 1975–June 1975.

Robert Lucas, Carnegie-Mellon University: visiting scholar, University of Chicago, June 1, 1974.

Basil J. Moore, Wesleyan University: visiting professor of economics, department of comparative social science, Universiti Sains Malaysia, Minden, Penang, June 1, 1974.

Daniel H. Newlon, State University of New York at Binghampton: staff associate in economics, Social Science Division, National Science Foundation, Aug. 1974.

Saul Pleeter, Indiana University: Brookings Economic Policy Fellow, U.S. Department of Labor, July 1, 1974.

Edward Prescott, Carnegie-Mellon University: Guggenheim Fellowship, Norwegian School of Economics and Business Administration, Norway, June 1, 1974.

Allan G. Pulsipher, Southern Illinois University: senior staff economist, Council of Economic Advisers, 1974–75.

James J. Rakowski, University of Notre Dame: senior lecturer, University of Dar Es Salaam, 1974–75.

Milton R. Russell, Southern Illinois University: senior staff economist, Council of Economics Advisers, 1974–75.

Todd Sandler, Arizona State University: visiting assistant professor of economics, State University of New York at Binghamton.

Abdelaleem M. Sharshar, Virginia Commonwealth University: senior adviser, Ministry of Commerce and Industry, Saudi Arabia, July 1974.

Robert Shelton, Arizona State University: Brookings Economic Policy Fellow, U.S. Water Resources Council, Washington.

Gordon W. Smith, Rice University: Deputy Assistant Secretary for International Affairs, U.S. Treasure Department, Washington, 1974–75.

George J. Stolnitz, Indiana University: chief demographer for Economic Commission of Europe, United Nations, July 1, 1974.

Ernst W. Stromsdorfer, Indiana University: director of evaluation for Assistant Secretary for Policy Evaluation and Research, U.S. Department of Labor, July 1, 1974.

Donald A. Wittman, University of California at Santa Cruz: University of Chicago, fall 1974.

Paul Wonnacott, University of Maryland: associate director, Division of International Finance, Board of Governors, Federal Reserve System. 1974–75.

Resignations

Bruce R. Beattie, Iowa State University: Texas A&M University, June 20, 1974.

David Cass, Carnegie-Mellon University: University of Pennsylvania, June 1, 1974.

Ronald Cerwonka, St. Bonaventure University: Providence College, Sept. 1974.

Flora D. Gill, University of California at Santa Cruz, July 1974.

Ronald E. Grieson, Queens College, City University of New York: Columbia University, June 1974.

Larry M. Hersh, Queens College, City University of New York: Hunter College, June 1974.

Michael Klass, University of Michigan: Charles River Associates, Cambrige, Mass., July 1, 1974.

James A. Largay III, Rice University: Georgia Institute of Technology, June 1, 1974.

Richard B. Mancke, University of Michigan: Tufts University, July 1, 1974.

Leo V. Mayer, Iowa State University: Library of Congress, July 1, 1974.

Kendrick W. Miller, University of California at Santa Cruz, July 1974.

William G. Murray, Iowa State University, June 1, 1974.

Leonard Rapping, Carnegie-Mellon University: University of Massachusetts, June 1, 1974.

Jati K. Sengupta, Iowa State University: Indian Institute of Management, Calcutta.

Donald C. Shoup, University of Michigan: School of Architecture and Urban Planning, University of California, Los Angeles, July 1974.

S. Richard Shumway, North Carolina State University, May 1974.

Irwin H. Silberman, Queens College, City University of New York, June 1974.

Lester D. Taylor, University of Michigan: University of Arizona, July 1, 1974.

T. Dudley Wallace, North Carolina State University, Aug. 1974.

Larry J. Wipf, Ohio State University: U.S. Bureau of Labor Statistics, Washington, Sept. 1974.

NOTE TO DEPARTMENTAL SECRETARIES AND EXECUTIVE OFFICERS

When sending information to the *Review* for inclusion in the Notes Section, please use the following style:

A. Please use the following categories:

1—Deaths
2—Retirements
3—Foreign Scholars (visiting the USA or Canada)
4—Promotions
5—Administrative Appointments
6—New Appointments
7—Leaves for Special Appointments (NOT Sabbaticals)
8—Resignations
9—Miscellaneous

B. Please give the name of the individual (SMITH, John W.), his present place of employment or enrollment: his new title (if any), his next place of employment (if known or if changed), and the date at which the change will occur.

C. Type each item on a separate 3×5 card, and please do not send public relations releases.

D. The closing dates for each issue are as follows: *March*, November 1; *June*, February 1; *September*, May 1; *December*, August 1.

This announcement supersedes and replaces a letter which was sent annually from the managing editor's office. All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

The following Statement of Ownership, Management, and Circulation is provided in accordance with the requirements, as contained in 39 U.S. Code 3685. The *American Economic Review* is owned, managed and published by the American Economic Association, a nonprofit educational organization, located at 1313 21st Avenue So., Nashville, Davidson County, Tennessee 37212. The Managing Editor is Professor George H. Borts, *American Economic Review*, Brown University, Robinson Hall, Providence, R.I. 02912. During the preceding 12 months the average number of copies printed for each issue was 27,900, the average paid circulation, 24,620; the average free distribution, 309; the average number of copies distributed, 24,929. Corresponding figures for the last issue before filing: 27,500 total number copies printed; 24,663 total paid circulation; 309 free copies distributed; 24,972 total distribution.

SEVENTY-FIRST LIST OF DOCTORAL DISSERTATIONS IN POLITICAL ECONOMY IN AMERICAN UNIVERSITIES AND COLLEGES

The present list specifies doctoral degrees conferred during the academic year terminating June 1974. Abstracts will no longer be printed, as they are published by University Microfilms, Ann Arbor.

General Economics; including Economic Theory, History of Thought, Methodology, Economic History, and Economic Systems

CARSON E. AGNEW, Ph.D. Stanford 1974. The dynamic control of congestion-prone systems through pricing.

JAMES M. ARNOTT, Ph.D. Georgia 1973. The relationship between labor market conditions and changes in the money supply.

ZAKARIA BASHA, Ph.D. Kansas 1973. Some economic welfare aspects of the theory of taxation.

WILLIAM M. BATESON, Ph.D. Wisconsin 1974. The agrarian system of the Upper Gangetic Plain, 1870–1910.

JEAN-PASCAL BENASSY, Ph.D. California (Berkeley) 1974. Disequilibrium theory.

JEROME T. BENTLEY, Ph.D. Pittsburgh 1974. The effects of Standard's integration into transportation on the structure, location, and efficient use of transportation in the petroleum refining industry: 1872–1885.

JACOB BENUS, Ph.D. Michigan 1974. Income instability: An empirical analysis.

RONALD E. BIRD, Ph.D. North Carolina 1974. A history and clarification of economic rent.

AMITAVA BOSE, Ph.D. Rochester 1974. Pareto-optimum consumption loans with and without production.

DARCY A. CARR, Ph.D. Pennsylvania State 1974. Factors influencing migration from the United Kingdom to the United States, 1860–1914.

ERIC T. CHESTER, Ph.D. Michigan 1973. Alienation and consumption: An extension of life cycle planning.

VICTOR CHOLEWICKI, Ph.D. Brown 1974. Effects of taste and technological change on the stability of competitive equilibrium.

LEONARD DENTE, Ph.D. New York 1974. Veblen's theory of social change.

ARTHUR T. DENZAU, Ph.D. Washington (St. Louis) 1973. Majority voting and general economic equilibrium: A synthesis.

MALCOLM DOLE, JR., Ph.D. California (Los Angeles) 1974. An economic theory of bureaucracy.

LEONARD T. ELZIE, Ph.D. Florida State 1974. The evolution of the scarcity concept.

JOHN F. ERMISCH, Ph.D. Kansas 1973. A quantitative analysis of the differential growth of American cities in the late nineteenth century.

ROGER L. FAITH, Ph.D. California (Los Angeles) 1973. A general equilibrium model of non-competitive interdependence.

MURRAY FESHBACH, Ph.D. American 1974. The structure of supply and demand for manpower in the *USSR*, 1850–1880.

JAMES M. FESMIRE, Ph.D. Florida 1973. Micro and macro aspects of public goods supply.

JORAM FISCHER, Ph.D. Massachusetts Institute of Technology 1974. General equilibrium and government action under uncertainty.

ROGER N. FOLSOM, Ph.D. Claremont School 1974. Linear constant coefficient difference or differential equations for economic models, 1973.

VALERIANO GARCIA, Ph.D. Chicago 1973. Argentine economic history, 1946–1970.

ROBERT W. GILMER III, Ph.D. Texas (Austin) 1973. Public financing of equal educational opportunity in a fiscal federalism.

PHILIP E. GRAVES, Ph.D. Northwestern 1973. The increasing relative risk aversion hypothesis.

JAMES P. GUNNING, JR., Ph.D. Virginia Polytechnic Institute 1974. An economic approach to social interaction and institutions.

KARMA G. HADJIMICHALAKIS, Ph.D. Rochester 1974. Consumption externalities in general equilibrium theory.

RICHARD B. HARRIFF, Ph.D. California (San Diego) 1973. Economic theory and distribution lags.

DANNIE HARRISON, Ph.D. Southern Illinois 1974. The economics of John Atkinson Hobson.

TATSUO HATTA, Ph.D. Johns Hopkins 1973. A theory of piecemeal policy recommendations.

MIHALIS HATZIPROKOPIOU, Ph.D. State University of New York (Binghamton) 1974. On the transformation problem.

MARTIN F. HELLWIG, Ph.D. Massachusetts Institute of Technology 1974. Sequential models in economic dynamics.

DEREK C. JONES, Ph.D. Cornell 1974. The economics of British producer cooperatives.

EDWARD A. KASCHINS, Ph.D. Iowa 1973. Karl Marx's theory of economic crises.

BARRY P. KEATING, Ph.D. Notre Dame 1974. Behavioral theory applied to the nonprofit firm: A theoretical and empirical analysis of credit unions.

JOHN C. KELLY, Ph.D. Boston College 1974. Patterns in Soviet industrial location: Case studies of the cement, granulated sugar, and electric power industries.

JOHN F. KENNAN, Ph.D. Northwestern 1973. Theories of factor demand of a competitive firm with adjustment costs and stochastic prices.

TETSUYA KISHIMOTO, Ph.D. Rochester 1974. Three essays on externalities.

ROBERT W. KLEPPER, Ph.D. Chicago 1973. The economic bases for agrarian protest movements in the United States, 1870–1900.

SHO-ICHIRO KUSUMOTO, Ph.D. Rochester 1974. The Le Chatelier-Samuelson principle in analytical economics.

DANIEL LEONARD, Ph.D. Illinois 1974. Competitive equilibrium and welfare maximization in an economy containing public nuisances.

VIVIEN LEVY-GARBOUA, Ph.D. Harvard 1974. Efficiency and optimality in macroeconomic policy: Two essays.

JOHN P. LEWIS, Ph.D. Ohio State 1974. The economics of intercommunications in the *USSR:* The postal service and telephone and telegraph systems.

CRAIG V. MCDONOUGH, Ph.D. Illinois 1973. Theory of economic imperialism.

JOSEPH MACIARIELLO, Ph.D. New York 1973. A comparative-dynamic procedure for benefit-cost analysis.

ARUP K. MALLIK, Ph.D. Rochester 1974. Structure of time-preference and the theory of long-run equilibrium.

MARILYN E. MANSER, Ph.D. Wisconsin (Madison) 1974. Estimating consumer preferences and cost of living indexes for *U.S.* meat and produce, 1947–1971.

DAVID MARCINKO, Ph.D. Boston College 1973. Centralized versus decentralized production: A comparative study of production functions for Polish state and private agriculture.

NABUHIKO MASUDA, Ph.D. Case Western 1974. Optimal pricing by dominant firms under uncertainty.

EDWARD MATLUCK, Ph.D. New York 1974. Factor demand functions in the *U.S.* manufacturing and nonmanufacturing industries: A dynamic approach.

AURELIO A. MATTEI, Ph.D. Johns Hopkins 1973. An intertemporal model of consumer behavior.

MICHAEL MEEROPOL, Ph.D. Wisconsin (Madison) 1973. On the origins of the Chilean nitrate enclave.

SWADNA MUKHOPADHYAY, Ph.D. Massachusetts Institute of Technology 1974. Studies in production functions.

MARTIN MURPHY, Ph.D. Boston College 1974. The role of households in economic decision making in a planned socialist economy: Poland.

DAVID M. NIENHAUS, Ph.D. Northwestern 1973. Profit-maximizing sales decisions for a production-to-order firm.

RAOUL ORESKOVIC, Ph.D. New York 1973. Power and neoclassical economic theory.

WILLIAM P. ORZECHOWSKI, Ph.D. Virginia Polytechnic Institute 1974. Economic models of bureaucracy: Survey, extensions, and evidence.

THEODORE J. OSBORNE, Ph.D. Brown 1974. Factor substitution and scale effects.

ALLEN M. PARKMAN, Ph.D. California (Los Angeles) 1973. Simultaneous legal and illegal demand.

ROBERT W. PULSINELLI, Ph.D. Rutgers 1974. Nonincome determinants of the propensity to consume: A principal components approach.

JOSEPH D. REID, JR., Ph.D. Chicago 1974. Sharecropping as an understandable market response: The postbellum South.

THOMAS P. REINWALD, Ph.D. Illinois 1973. Antecedents of the Chamberlin-Robinson analysis of noncompetitive markets.

LAURA J. RUBIN, Ph.D. Pennsylvania 1973. A theory of franchising.

ALVIN C. RUPPERT, Ph.D. California (Riverside) 1974. Irving Fisher's theory of capital and interest.

JOSE A. SCHEINKMAN, Ph.D. Rochester 1974. On optimal steady states of *n*-sector growth model when utility is discounted.

WILLIAM W. SHARKEY, Ph.D. Chicago 1973. A study of markets involving increasing returns and uncertain demand.

STEVEN M. SHAVELL, Ph.D. Massachusetts Institute of Technology 1974. Essays in economic theory.

JAE KANG SHIM, Ph.D. California (Berkeley) 1973. Spatial equilibrium analysis of the *U.S.* lumber market: An application of quadratic programming.

DEMETRIOS N. SKIOTIS, Ph.D. Pennsylvania 1973. An econometric model of the Byzantine Empire.

RICHARD K. SMITH, Ph.D. Wisconsin (Madison) 1973. The economics of education and discrimination in the *U.S.* South: 1890–1910.

CHARLES W. STAHL, Ph.D. California (Santa Barbara) 1974. A spatial theory of monopsonistic exploitation and its implications.

JON G. SUTINEN, Ph.D. Washington 1973. An economic theory of share contracting.

CAROL TAYLOR, Ph.D. Michigan 1974. The demand for money and uncertainty.

STEPHEN J. THOMPSON, Ph.D. Illinois 1973. Mormon economic thought: 1830–1900.

PETER G. TOMLINSON, Ph.D. Johns Hopkins 1973. Market mechanisms and the production of public goods.

HAL R. VARIAN, Ph.D. California (Berkeley) 1974. Equity, envy, and efficiency.

VINOD VYASULU, Ph.D. Florida 1974. Economic analysis and Baran development.

ROBERT C. WEST, Ph.D. Northwestern 1974. Theoretical and structural aspects of banking reform, 1863–1923: Real bills and the question of organization.

JAMES E. WILEN, Ph.D. California (Riverside) 1973. Issues in environmental decision analysis: Uncertainty, irreversibilities, and option value.

LONNY L. WILSON, Ph.D. Iowa 1973. Intercity wage and cost-of-living differentials in the United States, 1889–1939.

JAMES L. WOLFENBARGER, Ph.D. Tennessee 1974. Investment allocation and financing in Yugoslavia in the 1960's.

DEVON YOHO, Ph.D. Missouri 1974. Discussion groups in introductory economics: The effects of out-of-class student interaction.

Economic Growth and Development; including Economic Planning Theory and Policy, Economic Fluctuations and Forecasting

KATHRYN I. ABBASSI, Ph.D. California (Berkeley) 1973. Distributional consequences of growth.

MANABENDRA ADHIKARY, Ph.D. Indiana 1973. Economic growth in India, 1861–1961: Conjectural estimates on a controversial issue.

EDWARD ALBAN, Ph.D. Georgia 1973. The relationship between equilibrium gaps, potential gaps, and the inflation rate: A forecasting model.

FAISAL S. AL-BASHIR, Ph.D. Arizona 1973. An econometric model of the Saudi Arabian economy: 1960–1970.

NICOS E. ALEXANDRAKIS, Ph.D. Kentucky 1973. Tourism as a leading sector in economic development: A case study of Greece.

MICHAEL J. APPLEGATE, Ph.D. Iowa State 1973. A multi-sectoral linear programming approach to measurement of the economic costs of income redistribution of Guatemala.

VEMURI A. AVADHANI, Ph.D. Nebraska 1973. Imports and capital formation in the less developed countries.

MUHAMMAD H. AWAD, Ph.D. Colorado 1974. Nationalization of the production stage of oil industry in the Middle East: Economic cost-benefit analysis.

HAMID BAGHSHOMALI, Ph.D. California (Los Angeles) 1973. Optimal economic planning with a structural break: A study of the impact of the steel industry on the Iranian economy.

MARY ANN R. BAILY, Ph.D. Massachusetts Institute of Technology 1974. Capital utilization in Kenya manufacturing.

CHAIN S. BARLA, Ph.D. Michigan State 1973. An analysis of cooperative agricultural credit institutions in India: A case study of the primary credit societies in Rajasthan.

ROGER L. BAUR, Ph.D. Ohio State 1974. Historical description of capital and technology changes at the farm level in four southern Brazil regions: 1960–1969.

DAVID C. BEEK, Ph.D. Pennsylvania 1973. A macroeconometric analysis of the impact of external resources on a developing economy: The case of India.

NEVILLE O. BEHARIE, Ph.D. Illinois (Urbana) 1973. Economy of Iran.

ABEL BELTRAN DEL RIO LOZANO, Ph.D. Pennsylvania 1973. A macroeconometric forecasting model for Mexico: Specification and simulations.

KARL M. H. BENNETT, Ph.D. McGill 1974. Industrialization by invitation: An examination of the Jamaican and Puerto Rican experience, 1950–67.

STEPHEN D. BIGGS, Ph.D. California (Berkeley) 1974. A multisector model for the agricultural socio-economic system: Purnea District, Bihar, India.

PETER E. BLOCH, Ph.D. California (Berkeley) 1974. The structure of nonagricultural earnings in Brazil.

SAWAI BOONMA, Ph.D. Claremont School 1974. Agricultural contributions to economic development: A study of postwar Thailand and Meiji Japan.

SHANNON R. BROWN, Ph.D. California (Berkeley) 1974. A comparative study of collective farming in Russia and China.

TRUONG QUANG CANH, Ph.D. Ohio State 1974. Income instability and consumption behavior: A study of Taiwanese farm households 1964–1970.

MANUEL J. CARVAJAL, Ph.D. Florida 1974. An economic analysis of the demographic aspects of human capital investment in Costa Rica.

DENNIS L. CHINN, Ph.D. California (Berkeley) 1974. Potential effects of income redistribution on economic growth constraints: Evidence from Taiwan and South Korea.

WILLIAM E. CLARK, Ph.D. Harvard 1974. Socialist development and public investment in Tanzania 1964–1973.

NEAL P. COHEN, Ph.D. Wisconsin (Madison) 1974. Structural change in a low income country: An econometric approach using the Republic of Kenya.

KENTON R. CORUM, Ph.D. California (Berkeley) 1973. The demand for human resources in agriculture in relation to government price policies in economic development: The case of India.

JERRY CROMWELL, Ph.D. Harvard 1974. Income inequalities, discrimination, and uneven capitalist development.

TUGRUL CUBUCKU, Ph.D. New York 1973. Sources of productivity in Turkish manufacturing, 1950–1968.

JOHN M. CUMMINS, Ph.D. Northwestern 1973. Cost over-runs in military procurement: An empirical analysis of entrepreneurial choice under risk.

MICHAEL D. CURLEY, Ph.D. Kentucky 1974. Investment behavior and institutional change in rural South Vietnam.

KEMAL DERVIS, Ph.D. Princeton 1973. Substitution, employment, and intertemporal equilibrium in a multi-sector planning model for Turkey.

ARNOLD K. E. DE SILVA, Ph.D. McGill 1974. Financial aspects of economic development: The Ceylonese experience.

LAWRENCE E. DE SIMONE, Ph.D. California (Berkeley) 1974. A short-run regional econometric forecasting model of residential construction activity: An application to California standard metropolitan statistical areas.

WALTER G. DONOVAN, Ph.D. Cornell 1974. Employment generation in agriculture: A study in Mandya District, South India.

MITSUO EZAKI, Ph.D. Harvard 1974. Quantitative study of Japan's economic growth, 1952–1980: An approach from the system of national accounts.

YIU-KWAN FAN, Ph.D. Wisconsin (Madison) 1974. A multisector adaptive model of economic development and rural-urban migration.

HELEN M. FENTON, Ph.D. Boston College 1973. Voluntary foreign aid: The economics of charity.

JAMES B. FITCH, Ph.D. Stanford 1974. Economic development in a minority enclave: The case of the Yakima Indian nation, Washington.

PREM C. GARG, Ph.D. Stanford 1974. Optimal economic growth with exhaustible resources.

ANIMESH GHOSHAL, Ph.D. Michigan 1974. Export commodities and economic development: The Liberian rubber industry.

NICHOLAS P. GLYTSOS, Ph.D. Syracuse 1974. Prices, wages, and unemployment in the urban sector of the less developed countries: An intertemporal cross-sectional econometric model.

DUANE B. GRADDY, Ph.D. Lehigh 1974. The empirics of the equilibrium balanced growth model.

HARVEY N. GRAM, Ph.D. Wisconsin (Madison) 1973. Homogeneous and heterogeneous capital models in the theory of growth and the pure theory of international trade.

HITIHAMY M. GUNASEKERA, Ph.D. California (Los Angeles) 1973. Foreign resources and economic development of Ceylon.

FAYEZ I. HABIB, Ph.D. Indiana 1973. Savings deter-

minants in low income countries: The case of Egypt 1946–65.

CLAUDIO L. S. HADDAD, Ph.D. Chicago 1974. Growth of Brazilian real output, 1900–1947.

OLIVER D. HART, Ph.D. Princeton 1974. Essays in the economics of uncertainty.

ARTHUR HOLCOMBE, Ph.D. New York 1973. The capacity of the U.N. system to promote increased output, income, and employment among traditional producers in rural areas.

HAJIME HORI, Ph.D. Brown 1974. Economic growth of a two-country world with capital mobility.

EDWARD HUDSON, Ph.D. Harvard 1973. Optimal growth policies.

DONALD N. IKE, Ph.D. California (Los Angeles) 1973. Agricultural excess population and the economic consideration for land reform in Nigeria.

KHOONTONG INTARATHIA, Ph.D. Illinois (Urbana) 1974. The impact of foreign investment on the industrialization of Thailand.

DESMOND JOLLY, Ph.D. Oregon 1973. Sectoral growth and employment in the Jamaican economy, 1959–1968.

AHMAD A. KADER, Ph.D. West Virginia 1974. The role of the oil export sector in the economic development of Iraq.

MIHSSEN M. KADHIM, Ph.D. Colorado 1974. The strategy of development planning and the absorptive capacity of the economy: A case study of Iraq.

ANTONIOS E. KARAM, Ph.D. Temple 1974. Economic dependence and economic growth: An analytical framework and empirical study.

DANIEL R. KAZMER, Ph.D. Massachusetts Institute of Technology 1974. The agricultural development of Siberia, 1890–1917.

VIRABHAI C. KHARADIA, Ph.D. Illinois (Urbana) 1973. Monetary phenomena and economic growth with special reference to India.

SUNG T. KIM, Ph.D. Catholic. Indicators of liquidity and depth servicing difficulties of developing countries: The short-run empirical analysis.

SOO K. KOH, Ph.D. New York 1973. Commodity composition of exports of manufactures from a developing country: The case of Republic of Korea.

SANG-KYUNG KWAK, Ph.D. State University of New York (Albany) 1974. A study of price and wage behavior in the *U.S.* manufacturing industries, 1964–71.

DAVID L. LANDES, Ph.D. Washington (St. Louis) 1973. Import substitution and the demand for labor in urban Chile, 1930–70.

STANLEY LAWSON, Ph.D. New York 1973. The Jamaican economy: Prospect and retrospect.

PAUL K. LIU, Ph.D. Michigan State 1973. Interactions between population growth and economic development in Taiwan.

CHARLES L. MACDONALD, Ph.D. California (Berkeley) 1973. Instrument/target relationships for static and dynamic fixed target policy models.

MAURICE MACDONALD, Ph.D. Michigan 1974. Income distribution and economics mobility in Taiwan.

ALEXANDER MACMILLAN, Ph.D. Queen's 1974. Cyclical and secular inflation: A market oriented approach.

ROBERTO B. M. MACEDO, Ph.D. Harvard 1974. Models of the demand for labor and the problem of labor absorption in the Brazilian manufacturing sector.

MICHAEL MARTIN, Ph.D. New York 1973. Incomes policy in the United Kingdom and Finland and their relevance to the United States.

FERNANDO J. MARTINEZ, Ph.D. California (Berkeley) 1973. An economic analysis of the Chilean fresh and canned fruit industries.

LAL N. MIRCHANDANI, Ph.D. Southern California 1974. A comparison of employment performance in the manufacturing sector of the less developed countries: Empirical tests of alternate hypotheses.

ISMET MUCEK, Ph.D. New York 1973. Exports and economic growth in primary producing countries: The Turkish experience since 1960.

JOSE T. MULLEADY, Ph.D. Iowa State 1973. Technological change: The case of corn production in the Argentine Pampas.

WINTHROP MUNRO, Ph.D. Michigan 1973. An economic study of maize marketing in Kenya: 1952–66.

LEOPOLD P. MUREITHI, Ph.D. Claremont School 1974. Employment, technology, and industrialization in Kenya: A study in development strategy.

WILFRED A. NDONGKO, Ph.D. Carleton 1973. The theory and practice of planning for economic development in a federal state: The case of Cameroon.

DESMOND A. NORTON, Ph.D. California (Berkeley) 1974. The macro stage in Irish planning, 1958–72: A study in the quantitative theory of economic policy.

GEOFFREY E. NUNN, Ph.D. Virginia 1973. Aggregate price expectations: An alternative approach.

MUDZIVIRI T. NZIRAMASANGA, Ph.D. Stanford 1973. The copper export sector and the Zambian economy.

JOSEPH ODAMA, Ph.D. Missouri 1973. Adaptive control rules for economic planning: An application to the Nigerian economy.

REUBEN N. OGBUDINKRA, Ph.D. Catholic. A quantitative and qualitative evaluation of public investment criteria in the West African countries: National development plans.

GUNTAC OZLER, Ph.D. New York 1973. Optimal planning, prices, and the labor theory of value.

JONG-GOO PARK, Ph.D. Michigan State 1973. Choice of performance index for the optimal control of macroeconomic systems.

DOUGLAS K. PEARCE, Ph.D. Wisconsin (Madison) 1974. The formation and economic effects of price expectations.

IZZET PEKARUN, Ph.D. New York 1973. Excess capacity in Turkish manufacturing industry.

ALVARO POSADA, Ph.D. Michigan State 1974. A simulation analysis of policies for the northern Colombia beef cattle industry.

VAMON RAO, Ph.D. Missouri 1974. Capital stock and income distribution in a growing economy: An empirical study of the United States and India.

THOMAS REICHMANN, Ph.D. Harvard 1973. Persistent inflation and macroeconomic equilibria: The case of Chile, 1960–69.

S. NICHOLAS SAMUEL, Ph.D. Michigan State 1973. Income redistribution through technological changes in agriculture: Implications for Sri Lanka's paddy subsector.

ALAN R. SHELLY, Ph.D. California (Berkeley) 1974. A study in disequilibrium macroeconomics.

NADIMINTI R. SIVASWAMY, Ph.D. Wisconsin (Madison) 1974. Employment potential of the Indian industrial sector.

RICHARD S. SMITH, Ph.D. Texas (Austin) 1974. Diversification alternatives and inducements in the Colombian coffee industry.

JOHN N. STITZLEIN, Ph.D. Ohio State 1974. The economics of agricultural mechanization in southern Brazil.

TUNCAY M. SUNMAN, Ph.D. Rice 1974. Short-run effects of income distribution on some macroeconomic variables: The case of Turkey.

KENNETH G. SWANBERG, Ph.D. Cornell 1973. The potential impact on nutritional status of reducing marketing costs through marketing system manipulation in low income developing countries.

CLYDE G. SWENSON, Ph.D. Michigan State 1973. The effect of increases in rice production on employment and income distribution in Thanjavur District, South India.

ISLAM R. TARAFDER, Ph.D. Michigan State 1973. Alternative low-life pump irrigation policies in Bangladesh: An economic and financial analysis.

ASEFA TEFERRI, Ph.D. Michigan State 1973. Incomes policy for East Africa with special reference to Uganda.

GEORGE THOMA, Ph.D. California (Los Angeles) 1973. An annual econometric model for Cyprus, 1958–1971.

ROBERT L. THOMPSON, Ph.D. Purdue 1974. The metaproduction function for Brazilian agriculture: An analysis of productivity and other aspects of agricultural growth.

MARGO M. THORNING, Ph.D. Georgia 1974. An empirical investigation of money, real output, and prices over the business cycle, 1920–70.

MARY A. THROUP, Ph.D. State University of New York (Binghamton) 1974. Capital and output in the Mexican economy, 1939–67.

VU THUY, Ph.D. Michigan State 1974. A macroeconomic model for the evaluation of the irrigation impacts of the Pa Mong Project on the economy of South Vietnam.

WILLIAM D. TOAL, Ph.D. Notre Dame 1974. An aggregative expectational model applied to southeastern metropolitan areas.

JORGE A. TORRES, Ph.D. California (Berkeley) 1973. Input-output models and structural analysis of the Peruvian economy.

CHYAU TUAN, Ph.D. Ohio State 1973. Determinants of financial savings in farmers associations in Taiwan 1960–70.

JOHN G. VOYATZIS, Ph.D. Pennsylvania 1973. International comparisons of investment and consumption sector size.

THOMAS W. WARKE, Ph.D. Princeton 1973. The theory and measurement of learning from experience and of externalities in developing countries.

RICHARD C. WEBB, Ph.D. Harvard 1974. Government policy and the distribution of income in Peru, 1963–73.

STEPHEN F. WEBER, Ph.D. Wisconsin (Madison) 1974. Project evaluation of the consolidation of small mines in developing countries: The case of Chile.

ROBERT L. WELCH, Ph.D. Northwestern 1974. Iterative procedures for planning.

CHAIYAWAT WIBULSWASDI, Ph.D. Massachusetts Institute of Technology 1974. Thailand: A case study in open approach to foreign trade.

JOHN G. YOUNG, Ph.D. Harvard 1973. A cost-benefit analysis of the Area Development Incentives Act Program in the Atlantic Provinces.

JUNE P. YOUNG, Ph.D. Temple 1974. Land reform in South Korea.

Economic Statistics; including Econometric Methods, Economic and Social Accounting

JOSEPH O. ARATA, Ph.D. Kansas State 1974. A measurement of technological change by the use of the aggregate production function.

LEONARD B. AUERBACH, Ph.D. California (Berkeley) 1973. A population growth model for control decisions in agricultural systems.

GILBERT W. BASSETT, JR., Ph.D. Michigan 1973. Some properties of the least absolute error estimator.

STEWART W. BORLAND, Ph.D. Washington 1974. Short-run forecasting with regional input-output tables.

JAMES M. BRUNDY, Ph.D. California (Berkeley) 1974. The relative efficiency of instrumental variables estimators for the linear simultaneous equations model.

DONG W. CHO, Ph.D. Illinois (Urbana) 1973. Risk function of the preliminary test estimator in a simultaneous regression equations model.

JAMES L. DIETZ, California (Riverside) 1974. Optimal control of individual human capital investment.

ROBERT H. DUGGER III, Ph.D. North Carolina 1974. An application of bounded nonparametric estimating functions to the analysis of bank cost and production functions.

KENNETH M. GAVER, Ph.D. Carnegie-Mellon 1974. Comparing multiple equation models.

VINCENT J. GERACI, Ph.D. Wisconsin (Madison) 1974. Simultaneous equation models with measurement error.

DAVID K. GUILKEY, Ph.D. North Carolina 1973. Estimation of systems of regression equations with first-order vector autoregressive errors.

ROBERT S. GUTHRIE, Ph.D. Indiana 1973. A Monte Carlo study of the small sample properties of various estimators of distributed lag models.

DAVID H. HARRINGTON, Ph.D. Purdue 1973. Quadratic input-output analysis: Methodology for empirical general equilibrium models.

GEORG HASENKAMP, Ph.D. Wisconsin (Madison) 1973. Specification and estimation of multiple-output production functions.

MOHAMED HOSNY, Ph.D. North Carolina 1973. A macroeconometric model of the postwar *U.S.* economy: Estimation and simulation.

HOWARD J. HOWE, Ph.D. Pennsylvania 1974. Estimation of the linear and quadratic expenditure systems: A cross-section case for Colombia.

CAROLINA J. KAU, Ph.D. Missouri 1974. An econometric model of investment behavior in Missouri economy.

CHARLES F. KEITHAHN, Ph.D. Iowa State 1973. Estimation of the consumption function using the Almon technique.

LLOYD R. KENWARD, Ph.D. Queen's 1974. On a test for randomness in dynamic econometric autocorrelated error.

HEUNG R. KIM, Ph.D. Tennessee 1973. A study of saving of *U.S.* farm operators, 1946–70.

HYO KOO LEE, Ph.D. Northwestern 1973. An econometric model of Korea: 1958–70.

ANTHONY C. LEWIS, Ph.D. Purdue 1973. Experiments in the implementation of a convex programming algorithm.

MALCOLM A. LINDSAY, Ph.D. Wisconsin (Madison) 1974. A recursive programming model of the development of *U.S.* petroleum refining capacity: 1955–73.

PETER D. LOEB, Ph.D. Rutgers 1974. Specification errors and econometric models.

KEITH R. MCLAREN, Ph.D. Northwestern 1973. Spectral analysis and economic theory.

JAMES M. MALCOMSON, Ph.D. Harvard 1973. Production theory, replacement, and the utilization of capital equipment.

KHAJA H. MOINUDDIN, Ph.D. Hawaii 1974. An econometric model of India: A policy oriented approach.

CLAUDE MONTMARQUETTE, Ph.D. Chicago 1973. A model of inventory holdings with empirical application to Canadian manufacturing industries.

RICHARD P. O'NEILL, D.B.A. Maryland 1973. Trans-

formation methods for column generation procedures in non-linear programming.

PANAGIOTIS PAPKYRIAZIS, Ph.D. California (San Diego) 1974. Optimal design in econometric models.

JOHN B. PENSON, JR., Ph.D. Illinois 1973. An aggregative income and wealth model for the *U.S.* farm sector: Its description and application to policy analysis.

PAUL RAPPOPORT, Ph.D. Ohio State 1974. Least squares and its alternatives in the estimation of dynamic economic models.

CHAISUNG ROH, Ph.D. Illinois (Urbana) 1973. Import fluctuations and demand pressure in the United States: A disaggregated and distributed lag model.

ROMAN I. SENKIW, Ph.D. Virginia 1974. The growth of industrial production in Ukraine, 1945–71.

CLAIRE STARRY, Ph.D. Washington 1973. Some general equilibrium results for economies with producer nonconvexities.

EDWARD A. STOHR, Ph.D. California (Berkeley) 1973. Design of information systems for economic activities.

BRIAN P. SULLIVAN, Ph.D. North Carolina 1974. New estimates of the translog production function in *U.S.* manufacturing.

HOVAV TALPAZ, Ph.D. Michigan State 1973. Dynamic decision models in the hog-pork subsector production process.

KARAM C. TANEJA, Ph.D. Cincinnati 1974. Regional aspects of growth in output and productivity among *U.S.* manufacturing industries, 1958–69.

LLOYD D. TEIGEN, Ph.D. Michigan State 1973. Costs, loss, and forecasting errors: An evaluation of models for beef prices.

KIMIO UNO, Ph.D. Illinois (Urbana) 1973. Limits of Japanese growth: An econometric study.

ITZHAK VENEZIA, Ph.D. California (Berkeley) 1974. The optimal revision of information.

YIN-KANN WEN, Ph.D. Johns Hopkins 1974. An econometric study of the commercial paper market in the United States.

Monetary and Fiscal Theory, Policy, and Institutions

JAMES M. ARNOTT, Ph.D. Georgia 1973. An empirical investigation of money supply changes and the labor market.

MARTIN BALCI, Ph.D. New York 1973. An econometrics study of monetary sector of Turkey: 1948–70.

ROBERT J. BALL, Ph.D. Pennsylvania 1973. Inflation and the theory of money.

GARY L. BENJAMIN, Ph.D. Illinois 1973. An analysis of the structure and functions of correspondent banking: Implications for financing agriculture in Illinois.

JOHN N. BENSON, Ph.D. Queen's 1974. Cash and short-term liquid asset management in Canadian chartered banks, 1962–68.

JAMES M. BICKLEY, Ph.D. North Carolina 1973. An economic analysis of the North Carolina local-option sales tax referenda.

GREGORY E. BOCZAR, Ph.D. Florida 1974. The analytics of multibank holding company behavior.

PAUL W. BOLTZ, Ph.D. Illinois (Urbana) 1974. The economic impact of a statewide property tax in Illinois.

ROGER E. BRINNER, Ph.D. Harvard 1973. Neoclassical wage dynamics and involuntary development.

WINSTON R. CHIONG, Ph.D. Boston College 1974. A comparative study of the financial sectors of Australia and Ceylon.

NURUN N. CHOUDHRY, Ph.D. California (Berkeley) 1974. Fiscal, commercial and monetary policies in Pakistan (a methodological analysis of their effects).

NICOLA CINOSI, Ph.D. Fordham 1974. Luigi Einaudi's contribution to public finance.

JEFF R. CLARK, Ph.D. Virginia Polytechnic Institute 1974. Structures and incentives in the public provision: A comparative analysis of finance between community college systems of Virginia and North Carolina.

CLAUDIO R. CONTADOR, Ph.D. Chicago 1973. Money, inflation and the stock market: The Brazilian case.

DONNIE L. DANIEL, Ph.D. Illinois (Urbana) 1973. An analysis of the effects of market structure on performance of medium-sized banks.

JAMES M. DEAN, Ph.D. Virginia Polytechnic Institute 1974. Tax adjustments for inflation; with special emphasis on Canada.

CLARENCE R. DEITSCH, Ph.D. New Hampshire 1974. Economic stability and social security.

FOREST DENMAN, Ph.D. Georgia State 1973. Currency drain in the postwar United States.

WILLIAM M. DUGGER, Ph.D. Texas (Austin) 1974. Federal Reserve open market strategy in transition.

WALTER EBANKS, Ph.D. New York 1973. Government employment and post-World War II economic actuations: A study of the stabilizing effects of government employment.

JOHN H. ENNS, Ph.D. California (Los Angeles) 1973. The impact of federal grants-in-aid on state highway expenditures and revenues: An econometric study.

JOHN D. FERGUSON, Ph.D. Brown 1974. Estimation of macro-economic models: A study in the timing of monetary policy.

THOMAS K. FITZGERALD, Ph.D. Rutgers 1974. The growth of state government expenditures in the United States: Interpretations based on the appraisal of alternative hypotheses.

ROY FLORES, Ph.D. Iowa State 1973. The regional differential impact of monetary policy.

MARILYN R. FLOWERS, Ph.D. Virginia Polytechnic Institute 1974. Provision of law enforcement in a federal system.

HUGH B. GARNETT, JR., Ph.D. California (Berkeley) 1974. The general equilibrium incidence and effects of a value-added tax of the consumption type.

ABDUL GHAFUR, Ph.D. Iowa 1973. Dynamic differential tax incidence.

WILLIAM B. GREEN, Ph.D. Louisiana State 1974. An analytical and empirical examination of monetary multipliers in single equation macroeconomic models.

RAOUF S. HANNA, Ph.D. Indiana 1973. Credit market lag in the effects of monetary policy multivariate time and frequency domain analyses.

HELEN M. HARDY, Ph.D. McMaster 1973. The effect of federal grants on provincial expenditure and revenue decisions: Ontario and New Brunswick compared.

ARTHUR M. HAVENNER, Ph.D. Michigan State 1973. A computerized sensitivity analysis of selected assumptions in an annual econometric model of the *U.S.* economy: The Klein-Goldberger.

WAYNE A. HAYENGA, Ph.D. Michigan State 1973. The effects of bank mergers on financial services available to rural Michigan residents.

DONALD C. HINES, Ph.D. Kansas State 1974. Public primary and secondary education costs in Kansas with implications for revenue generation and allocation.

JOSEPH W. HOEBLER, Ph.D. Ohio State 1974. Monetary policy indicators and targets: Some issues.

JAMES A. HOEVEN, Ph.D. Colorado State 1973. The adequacy of Colorado's state-local tax structure.

ROBERT D. HOLLINGER, Ph.D. Kansas State 1974. Borrowing of Euro-dollars by *U.S.* commercial banks from their foreign branch banks and *U.S.* monetary control, 1964–72.

EDWARD T. HOWE, Ph.D. State University of New York (Albany) 1973. New York State Public School Districts capital replacement needs and financing.

MUHAMMAD HUSSAIN, Ph.D. Kansas State 1974. The impact of some monetary and fiscal variables on economic activity in Pakistan.

DOUGLAS N. KANE, Ph.D. Illinois (Urbana) 1973. State-level fiscal relationships.

JAMES B. KEHR, Ph.D. Ohio State 1974. A partial analysis of the factors that affect member bank reserves.

DONG-KUN KIM, Ph.D. Georgia 1973. An economic model of public expenditures on higher education: A study of demand and supply analysis of public goods.

TAEHO KIM, Ph.D. Colorado 1973. The money supply process under uncertainty.

JOHN J. KORBEL, Ph.D. Colorado 1974. The Colorado tax burden under alternative income concepts and tax shifting assumptions: A case study for fiscal year 1972.

HELEN LADD, Ph.D. Harvard 1974. Local public enterprise and the composition of the local property tax base.

DANIEL E. LAUFENBERG, Ph.D. Iowa State 1973. An econometric model of conventional mortgage credit flows and the channelled effects of monetary policy.

MARIO LECCISOTTI, Ph.D. California (Los Angeles) 1973. Taxation, asset prices, and the financial market under uncertainty.

ANTONIO C. B. LEMGRUBER, Ph.D. Virginia 1974. A study of the accelerationist theory of inflation.

SHARON G. LEVIN, Ph.D. Michigan 1973. The incidence of differential rates of property taxation on the business firm.

LUNG-HO LIN, Ph.D. Notre Dame 1974. A feedback control mechanism for money supply.

HENRI R. LORIE, Ph.D. Northwestern 1973. The transactions demand for money in a costly exchange economy.

JAMES R. LOTHIAN, Ph.D. Chicago 1973. The demand for high-powered money.

JAMES R. MCCABE, Ph.D. Missouri 1974. Portfolio approach to the term structure of interest rates.

DANIEL E. MCCARTY III, Ph.D. Georgia State 1973. An empirical investigation of the manufacturing sector's demand for money.

PRESTON J. MILLER, Ph.D. Minnesota 1972. The transactions demand for money in a three-asset economy.

JOSEPH P. MULHOLLAND, Ph.D. Washington (St. Louis) 1973. An empirical analysis of takeover activity: 1967–69.

ALICIA H. MUNNELL, Ph.D. Harvard 1973. The impact of social security on personal saving.

MICHAEL MUSSA, Ph.D. Chicago 1974. A Metzleric model of macro-economic dynamics.

GLENN L. NELSON, Ph.D. Michigan State 1973. The financing of higher education with specific application to Michigan.

EIJI NEZU, Ph.D. Pennsylvania 1973. Monetary and fiscal policy in a growing economy: An analysis of optimal monetary and fiscal policy, monetary and fiscal neutrality, and capital intensiveness.

THOMAS H. NICKELS, Ph.D. Illinois (Urbana) 1974. Intergovernmental grants: The role of the states as intermediaries.

JOSEPH NOLAN, Ph.D. New York 1973. The changing concept of the prime rate: An analysis and evaluation of New York banks' experience, 1968–72.

FREDERICK W. PARKHURST, JR., J.S.D. New York. The social utility of property taxation: Economics, law, and public policy.

PAUL R. PORTNEY, Ph.D. Northwestern 1973. Congressional delays and fiscal policy: A simulation of the costs and effects.

KATHLEEN A. POWERS, Ph.D. Houston 1973. The primary market for negotiable certificates of deposit: 1967–71.

DOUGLAS D. PURVIS, Ph.D. Chicago 1973. Approaches to the theory of asset behavior and macro-economic dynamics.

MARY J. RACSTER, Ph.D. Illinois (Urbana) 1973. The state-local income tax structure in Ohio.

IRA E. RASKIN, Ph.D. American 1973. Differential cost effectiveness of state and local spending: Implications for federal grants system.

RONALD R. REIBER, Ph.D. Arizona 1974. Directly placed commercial paper as an offset to *ex ante* monetary policy, 1959–70.

WATKINS L. RIBBLE, JR., Ph.D. Massachusetts Institute of Technology 1974. Portfolio behavior of *U.S.* life insurance companies.

THOMAS R. ROBINSON, Ph.D. George Washington 1974. Price-level expectations and the term structure of interest rates.

ARTHUR J. ROLNICK, Ph.D. Minnesota 1973. Evaluating the effectiveness of monetary reforms in reducing uncertainty about the impact of monetary policy.

JOHN RUTLEDGE, Ph.D. Virginia 1973. A monetarist model of inflationary expectations.

JACK L. RUTNER, Ph.D. Chicago 1974. Money in the antebellum economy: Its composition, relation to income and its determinants.

EFRAIM SADKA, Ph.D. Massachusetts Institute of Technology 1974. Income distribution, incentive effects, and optimal income taxation.

MICHAEL SAKELLIS, Ph.D. Clark 1974. The savings behavior of the household.

JOHN L. SCADDING, Ph.D. Chicago 1974. Monetary growth and aggregate savings.

OSVALDO SCHENONE, Ph.D. Chicago 1973. A dynamic analysis of taxation.

PAUL SCHNEIDERMAN, Ph.D. Clark 1974. The effect of credit market conditions on state and local government capital outlays.

LESLIE SMALL, Ph.D. Indiana 1973. An analysis of Mexican monetary management 1945–69.

SHELBY J. SMITH, Ph.D. Iowa State 1973. An analysis of the relationship between the mortgage credit supplied by financial intermediaries and residential housing activities.

WILLIAM SPAETH, Ph.D. Boston College 1974. Labor effort disincentives of negative income taxation.

JAMES M. STONE, Ph.D. Harvard 1973. An economic study of the securities industry.

MARIE E. SUSHKA, Ph.D. Georgetown 1974. An econometric model of the money market in the United States, 1823–59.

MICHAEL TENNENBAUM, Ph.D. California (Los Angeles) 1973. Financial intermediation, the demand for money and economic stability.

HARRIET G. TOLPIN, Ph.D. Boston College 1973. Optimal composition of the commercial bank balance sheet and the bank-customer relationship.

RICHARD L. TRIMBLE, Ph.D. Michigan State 1973. An

economic analysis of the effect of monetary policy on the beef industry.

WILLIAM B. TRUEHART, Ph.D. Claremont School 1973. The impact of real property versus land value taxation in Los Angeles County.

THOMAS VASQUEZ, Ph.D. Clark 1974. Differential taxation of income from capital in the United States, 1963–69: Analysis and management.

RODGER L. WEAVER, Ph.D. Utah. Planning, programming, budgeting system and alternative futures contingency budgets: A comparative theoretical analysis.

MARY E. H. WEBER, Ph.D. Utah. An approach for designing a general sales tax system: The Chilean case.

BYUNG-JOON WHANG, Ph.D. State University of New York (Binghamton) 1974. Inflation and economic growth: A dynamic analysis of deficit finance.

KENNETH J. WHITE, Ph.D. Wisconsin (Madison) 1973. The effect of bank credit cards on household financial decisions.

DELORES R. WRIGHT, Ph.D. Iowa State 1973. Two hypotheses of interest rate determination: An empirical test.

JAI-HOON YANG, Ph.D. California (Los Angeles) 1973. The determinants of membership in the Federal Reserve System.

MAHMOOD YOUSEFI, Ph.D. California (Santa Barbara) 1974. The relative efficacy of rules versus discretion in monetary management.

DAVID M. ZARNOCH, Ph.D. West Virginia 1973. Efficiency effects of a classified property tax: The West Virginia case.

International Economics

KINGSLEY AMOAKO, Ph.D. California (Berkeley) 1974. Balance-of-payments problems and exchange rate policy in a developing country: The Ghanaian experience.

SOBRAB BEHDAD, Ph.D. Michigan State 1973. National characteristics and the commodity composition of trade in manufacture goods.

PANAGIS A. BENETATOS, Ph.D. Southern Illinois 1974. The determinants of the sectoral distribution of private direct foreign investment: Greece, 1953–67.

DAVID BLOND, Ph.D. New York.1973. Foreign exchange rates, actual and expected, and the importer's decision process.

OSMAN BUBIK, Ph.D. California (Riverside) 1974. Cross-price elasticities in international trade.

JOHN D. CAMBON, Ph.D. California (San Diego) 1974. Foreign investment and trade between rich and poor countries: The Indo-American example.

ROBERT J. CARBAUGH, Ph.D. Colorado State 1974. Survey of the international monetary system.

LUCY A. CARDWELL, Ph.D. Massachusetts Institute of Technology 1974. Tariff reforms in a linear programming model of Chile.

EDWIN R. CARLISLE, Ph.D. Rutgers 1974. Structure of foreign sales: Exports and subsidiary production.

JOHN H. CHEH, Ph.D. Massachusetts Institute of Technology 1974. *U.S.* trade policy and short-run domestic adjustment costs.

HYUN SIK CHUNG, Ph.D. West Virginia 1973. A study of a model of international trade and investment between a developed and a developing country.

PAUL DE GRAUWE, Ph.D. Johns Hopkins 1973. The interaction of monetary policies in a group of European countries.

DOGAN DERELI, Ph.D. California (Berkeley) 1974. Trade policies and industrialization in Turkey.

JOHN K. DIETRICH, Ph.D. Michigan 1973. International capital transactions: A theoretical and econometric analysis.

JAMES A. DUNLEVY, Ph.D. Northwestern 1974. An econometric study of the short-run behavior of *U.S.* exports: A simultaneous equations approach.

MICHAEL E. EDO, Ph.D. Harvard 1974. Monetary instruments and the external balance.

STEPHEN ETTINGER, Ph.D. Michigan 1974. The economics of the customs union between Botswana, Lisotho, Swaziland, and South Africa.

DIETRICH K. FAUSTEN, Ph.D. Utah. The consistency of British balance-of-payment policies, 1949–67.

GARY L. FRENCH, Ph.D. Houston 1973. An empirical test of the Heckscher-Ohlin theory of international trade.

DAVID GALLO, Ph.D. Oregon 1974. The theory of effective protection, rates of protection to capital and resource allocation: A theoretical appraisal.

HANS A. GENBERG, Ph.D. Chicago 1974. Aspects of the monetary approach to balance-of-payments theory: An empirical study of Sweden.

NORMAN R. GIBSON, Ph.D. Massachusetts Institute of Technology 1974. The case for international money.

CLIFFORD L. GIONET, Ph.D. Florida 1974. A multiparty collective bargaining model as applied to international commodity agreements.

BENJAMIN B. GREENE, JR., Ph.D. Boston College 1974. Economic integration and import substituting industrialization in Central and Latin America.

DOMINIQUE HACHETTE, Ph.D. Chicago 1973. Revaluation of the escudo and distribution of income.

ARNOLD A. HEGGESTAD, Ph.D. Michigan State 1973. Market structure, profitability, and risk in the commercial banking industry.

ELHANAN HELPMAN, Ph.D. Harvard 1974. Macroeconomic policy and employment in an open economy with nontraded goods.

HUYNH HIEP, Ph.D. Michigan State 1974. Pollution and pollution controls: Their "real" and "financial" impacts on *U.S.* trade.

FERNANDO B. HOMEN DE MELO, Ph.D. North Carolina State 1973. An analysis of the world economy in 1980.

ROBERT C. KAMP, Ph.D. Carleton 1973. The impact of the balance of payments on the Canadian money supply (1962–1970).

MARYANN O. KEATING, Ph.D. Notre Dame 1974. The effect of international financial flows on the money supply process: A theoretical and econometric analysis of the *U.S.* experience in the 1960's.

ALFRED M. KRETZMANN, Ph.D. Colorado State 1973. Productivity change and *U.S.* trade.

WON-KEY KWON, Ph.D. Hawaii 1974. The balance of payments, money, and economic growth: A feedback mechanism.

MARTIN LANDSBERG, Ph.D. Wisconsin (Milwaukee) 1973. A simulation study of devaluation.

RALPH W. LANGE, Ph.D. Louisiana State 1974. An economic analysis of government exchange.

JOHN S. LAPP, Ph.D. Princeton 1974. Geographic variation of savings and loan deposit rates.

VINCENT L. LIN, Ph.D. Rutgers 1974. Trade preferences in the industrial development of Puerto Rico.

GEORGE LOGUSCH, Ph.D. New York 1973. An econometric model of Soviet foreign trade.

JAMES E. MCCARTHY, Ph.D. Fletcher School, Tufts 1974. Trade adjustment assistance: A case study of the shoe industry in Massachusetts.

JAMES F. MCCOLLUM, Ph.D. Rice 1974. Interest rates and inflation in Canada.

VINCENT J. MALANGA, Ph.D. Fordham 1974. An empirical study of fluctuations in the export proceeds of underdeveloped countries and their effect on economic development.

ROBERTO MARCHESINI, Ph.D. Texas (Austin) 1974. Impact of multinational corporations on domestic employment.

JAMES R. MARKUSEN, Ph.D. Boston College 1973. Cooperative and noncooperative control of international common property resources.

LEO MERORES, Ph.D. New York 1973. Haiti and the Dominican Republic: A case for economic integration.

ROBERT D. MONTGOMERY, Ph.D. Cornell 1974. The link between trade and labor absorption in rural Java: An input-output study of Jogjakarta.

JOHN H. MUTTI, Ph.D. Wisconsin (Madison) 1974. The employment and welfare effects of foreign trade policy.

CARMINE NAPPI, Ph.D. McGill 1974. Quantitative methods applied to the international and interprovincial trade of Quebec.

TITH NARANHKIRI, Ph.D. Colorado 1974. Economic performance and formulation of balance-of-payments policies within a fixed exchange rate system: Experience of Japan, 1961–70.

RUBEM D.F. NOVAES, Ph.D. Chicago 1974. Foreign investments in Brazil: An economic analysis.

HASAN OLGUN, Ph.D. Johns Hopkins 1973. The structure of protection and policies of industrialization in the Turkish manufacturing industries; 1960–70.

PHILIP OPHER, Ph.D. New York 1973. Policies related to foreign direct investments: A comparison between developed and developing countries.

EMILIOS PAGOULATOS, Ph.D. Iowa State 1973. The effects of agricultural and trade policies on European economic integration.

ROLANDO F. PELAEZ, Ph.D. Houston 1973. Exchange rate devaluation in a foreign exchange constrained economy: The Colombian case.

PHILIP PFAFF, Ph.D. Michigan State 1973. Forecasting the money stock.

EARL D. PHAUP, Ph.D. Virginia 1974. The discipline argument for fixed exchange rates.

WILFRIED E. PREWO, Ph.D. Johns Hopkins 1974. A multinational interindustry gravity trade model for the European Common Market.

JAMES W. REESE, Ph.D. Tennessee 1973. An analysis of bilateral trade of the OECD countries using the gravity model, 1959–69.

WILLIAM W. ROBERTS, Ph.D. California (San Diego) 1974. Monetary expansion, capital accumulation, and balance of trade.

CARLOS RODRIGUEZ, Ph.D. Chicago 1973. Aspects of international factor mobility.

DONALD ROUSSLANG, Ph.D. Oregon 1974. Short-term international capital flows and the effectiveness of monetary policy in an open economy.

EDMUND J. SHEEHEY, Ph.D. Michigan State 1974. Stabilization policy and protective discrimination in Argentina, 1967–71.

DICKSON K. SMITH, Ph.D. Utah. Foreign aid and economic development in Pakistan: The Ayub years (1958–1969).

JOSEPH E. SMOLIK, Ph.D. Illinois (Urbana) 1973. Domestically produced steel products to import competition: 1954–67.

BERNARD SNOY ET D'OPPVERS, Ph.D. Harvard 1974. The homogeneity of taxes on direct-investment income in the European economic community.

WILLIAM L. SPRINGER, Ph.D. Princeton 1973. Windfalls, temporary income taxes, and consumption behavior.

RICHARD J. STAKEM III, Ph.D. Southern Methodist 1974. The impact of domestic tariff protection on national resources: The case of Section 807.00 of the United States Tariff Schedule.

JOHN W. SUOMELA, Ph.D. Ohio State 1974. Optimal tariffs and optimal economic integration.

EDUARDO M. SUPLICY, Ph.D. Michigan State 1973. The effects of mini-devaluations on the Brazilian economy.

CHRISTOPHER A. TAYLOR, Ph.D. Princeton 1974. An "attributes" model of financial asset choice for mutual savings bank.

AHMET TEKINER, Ph.D. Boston College 1974. The effect of foreign resource transfers on Turkish economic development, 1952–65.

MASAHIDE TERAO, Ph.D. Illinois (Urbana) 1974. A bank portfolio selection theory for Japanese commercial banks.

MARIE C. THURSBY, Ph.D. North Carolina 1974. The resource reallocation costs of fixed and flexible exchange rates.

TUAN TON-THAT, Ph.D. Notre Dame 1974. Impact of linking new reserve creation to development assistance on international trade.

LUKE T. ULKITSET, Ph.D. West Virginia 1974. Exports of primary products as a catalyst to economic development: Thailand as a case study, 1950–70.

RUBENS VALENTINI, Ph.D. Purdue 1974. Technology and international trade in agricultural products: A test of some hypotheses.

JOHN J. VAN BELLE, Ph.D. Virginia 1974. An analysis of objective indicators of speculative activity under a system of flexible exchange rates.

JOHN WILSON, Ph.D. Pennsylvania 1974. Yet another econometric model of *U.S.* imports, 1958–71: Disaggregated by end-use commodity groups and region of origin.

BENNO WYMAR, Ph.D. Nebraska 1974. Regional cooperation for development: Iran, Pakistan, and Turkey.

AHMET YORUKOGLU, Ph.D. New York 1973. The effective rate of protection and developing countries: The Turkish case.

Business Administration; including Business Finance and Investment, Insurance, Marketing, and Accounting

ALEXANDER E. BERGMANN, Ph.D. California (Berkeley) 1973. The evolution of managerial attitudes toward organizational members and environmental responsibilities.

JOSEPH A. FEIGO, Ph.D. California (Riverside) 1973. Marketing resource allocation in the southern California electric utility industry.

ROGER H. GATES, Ph.D. Florida 1973. An examination of the determinants of residential electricity sales with special emphasis on the effects of advertising.

WILLIAM B. GILLESPIE, Ph.D. Florida 1974. The price impacts of reverse stock splits.

RAMESH C. GUPTA, Ph.D. California (Berkeley) 1973. Price-level accounting and the behavior of security prices.

MICHAEL HANSEN, Ph.D. California (Berkeley) 1974. Branch bank site selection: An empirical analysis of the importance of location and an approach to the capital budgeting problem.

JAMES R. HARRIS, Ph.D. Florida 1973. A model statute to regulate unfair advertising and sales practices in Alabama.

WILLIAM L. HARRISON, Ph.D. California (Berkeley) 1973. Management control in a large branch banking system: An empirical study.

AHARON HIBSHOOSH, Ph.D. California (Berkeley) 1974. A model of brand competition in attribute space: Toward a formal framework for marketing analysis of consumer behavior.

JOHN R. KELLY, Ph.D. Rice 1974. Dividend policy and the valuation of the firm.

ANDREW A. MITCHELL, Ph.D. California (Berkeley) 1973. An exploration into the relationship between benefit segmentation, multidimensional scaling and brand choice behavior using an expectancy theory framework.

DENIS P. NEILSON, Ph.D. California (Berkeley) 1974. An evaluation of computer instruction in undergraduate accounting curricula.

ANTONI P. PARÉS, Ph.D. Georgia 1972. Impact of the mutual funds industry on the stock market: The Spanish case, 1969–70.

ALLEN E. ROLLAND, D.B.A. George Washington 1974. Economic aspects of built-in test capability for modular electronic equipment.

LOUIS I. ROSEN, D.B.A. Maryland 1974. The legal process evolves a description of the accountant's disclosure obligation: An in-depth analysis of the Continental Vending Machine Corporation case.

ABDUL-KARIM T. SADIK, Ph.D. Florida 1973. The cost of equity capital for holding company subsidiaries.

ANDREW J. SENCHACK, JR., Ph.D. California (Los Angeles) 1973. The firm's optimal financial decisions: An integration of corporate financial theory under certainty.

JANET R. M. SMITH, Ph.D. California (Berkeley) 1974. A model for data selection policies in information system design: A theoretical approach.

WILLIAM S. STEWART, D.B.A. George Washington 1973. Providing repair support for surface ships: A generalized solution to a resource allocation problem.

COURTENAY C. STONE, Ph.D. California (Los Angeles) 1973. The impact of blue-sky regulations on the new issues market for corporate securities: Implications and evidence.

AHMET TEZEL, Ph.D. California (Berkeley) 1973. Optimal foreclosure policies.

MICHAEL J. VERTIGAN, Ph.D. California (Berkeley) 1974. Impact of alternative profit measures on selected macro-economic models.

RONALD J. WEBB, D.B.A. Maryland 1973. Organizational effectiveness in voluntary organizations: A study of the institutional church.

JEROLD L. ZIMMERMAN, Ph.D. California (Berkeley) 1974. Individual financial decision making in basic science: A normative and descriptive study.

Industrial Organization and Public Policy; including Economics of Technological Change, and Industry Studies

JEAN M. ADAMS, Ph.D. Illinois (Urbana) 1974. The pricing of doctor services.

WILLIAM J. ADAMS, Ph.D. Harvard 1973. Corporate power and profitability in the North Atlantic community.

ANN P. ANDERSON, Ph.D. Michigan 1974. The influence of manufacturing firms and industry mix on local communities.

CLAUS ARMIN, Ph.D. Tufts 1974. Optimal design and cost allocation for transportation networks.

ROBERT L. AYANIAN, Ph.D. California (Los Angeles) 1974. Advertising and rate of return.

LEE A. BAIR, Ph.D. Michigan State 1973. The effect of the introduction of medical assistance and medicare on the structure on the Michigan nursing home industry.

JOHN R. BALDWIN, Ph.D. Harvard 1973. A positive theory of regulation and the "public" corporation in the context of the Canadian air transport industry.

WILLIAM BARNETT, Ph.D. Michigan State 1974. Production functions for education: An empirical study.

GLEN BEESON, Ph.D. Pittsburgh 1973. Contributions of price theory and spatial economic theory toward defining the relevant market for antitrust.

ROBERT T. BENTON, Ph.D. West Virginia 1974. Municipal finance and rates of return for publicly owned electric utilities.

JEAN-THOMAS BERNARD, Ph.D. Pennsylvania 1973. The theory of the firm under rate of return regulation: A reexamination from the demand side.

LEIGH B. BOSKE, Ph.D. Pittsburgh 1973. The impact of managerial objectives, efficiency, and capacity level decisions on performance variables in the *U.S.* domestic trunk airline industry.

MYLES G. BOYLAN, Ph.D. Case Western 1974. The economics of changes in the scale of production in the *U.S.* iron and steel industry.

LEONARD BRONITSKY, Ph.D. Rensselaer 1973. The economics of construction mineral aggregates with an analysis of the industry in the greater New York metropolitan area.

BRIAN C. BRUSH, Ph.D. North Carolina 1973. Advertising intensity and market structure in consumer goods industries.

RONALD P. CERWONKA, Ph.D. Missouri 1974. An economic analysis of Nigeria frontier future air transportation facility requirements.

ROBERT CHATOV, Ph.D. California (Berkeley) 1973. The collapse of corporate financial standards regulation: A study of SEC-accountant interaction.

JOHN A. CHRISTIANSON, Ph.D. Wisconsin (Madison) 1974. The determinants of the rate of diffusion of new consumer products.

WILLIAM A. CLARKE, Ph.D. Rutgers 1974. Vertical integration in the aluminum industry.

DOUGLAS L. COCKS, Ph.D. Oklahoma State 1973. The impact of the 1962 drug amendments on R&D productivity in the ethical pharmaceutical industry.

DANIEL W. COLLINS, Ph.D. Nebraska 1974. The structure, conduct, and performance of the recording industry in the United States.

ROBERT L. CONN, Ph.D. Houston 1973. A theoretical and quantitative analysis of the causes and probable effects of pure conglomerate mergers, 1950–69.

LEON COURVILLE, Ph.D. Carnegie-Mellon 1974. Regulation and efficiency in the electric utility industry: An examination of the Averch-Johnson proposition.

CLAUDINE V. COX, Ph.D. Missouri 1974. Higher educational cost allocation: An economic synthesis.

JANE S. CROMARTIE, Ph.D. Florida 1974. The Federal Trade Commission's activities in the area of deceptive trade practices.

KENT A. CURRIE, Ph.D. Iowa 1973. Behavior of the regulated monopolist under demand uncertainty.

EDWARD V. DALEY, Ph.D. Missouri 1974. An economic evolution of alternative medical service systems in Columbia, Missouri.

BETTY A. DICHAUS, Ph.D. Missouri 1974. An economic evolution of Missouri automated radiology systems, MARS: A case study.

CARLTON L. DUDLEY, JR., Ph.D. California (Berkeley) 1974. A theoretical financial analysis of the long-term subsidy value of the regional development incentives program in Canada, and a related paper on corporate investment theory.

DAVID DUNLOP, Ph.D. Michigan State 1973. The economics of Uganda's health service system: Implications for health and economic planning.

VERNON L. FAHLE, Ph.D. Pittsburgh 1973. Ordinal measurement of manufactures locational sensitivity to transport costs.

JAMES M. FALVEY, Ph.D. Michigan State 1973. The adjustment of electric power generation to alternative SO_2 standards: A case study in environmental protection regulation.

LEE FETTNER, Ph.D. New York 1974. The American maritime industry since 1928: Productivity and other factors in its decline.

ALICE F. GERSTER, Ph.D. Boston College 1973. The change in social welfare from deregulation: The case of the natural gas industry.

HERBERT E. GISHLICK, Ph.D. Pennsylvania 1973. Plant location in manufacturing: A testing of some hypotheses.

GAYLON E. GREER, Ph.D. Colorado 1974. Risk, return, and efficiency in the market for real property.

WAYNE D. HALES, Ph.D. Oklahoma State 1974. An analysis of the effects of competitive markets on hospital costs in Oklahoma.

WILLIAM D. JACKSON, Ph.D. North Carolina 1974. Commercial bank regulation, structure, and performance.

M. JAVAD KHALILZADEH-SHIRAZI, Ph.D. Harvard 1973. Market structure and allocative efficiency: Studies in *U.K.* and *U.S.* manufacturing industries.

ROGER KOENKER, Ph.D. Michigan 1974. The estimation of input demand functions and the relative economic efficiency of regulated trucking firms.

TILMON KREILING, JR., Ph.D. Stanford 1974. Empirical estimation of *U.S.* demand for meat products.

JEROME K. LAURENT, Ph.D. Indiana 1973. The development of harbors, waterborne shipping, and commerce at six Wisconsin ports on Lake Michigan through 1910.

MANFRED H. LEDFORD, Ph.D. Kentucky 1973. The retail distilled spirits industry in Kentucky.

STANFORD LEVIN, Ph.D. Michigan 1974. The impact of entry on large *U.S.* industrial corporations.

LYCURGUS L. LIAROPOULOS, Ph.D. Michigan State 1973. A measure of productive efficiency with application in incentive reimbursement for hospital care.

STEWART L. LONG, Ph.D. Illinois (Urbana) 1974. The history of the DuMont television network 1946–55, "The Economics of a Fourth Network."

PAUL B. MANCHESTER, Ph.D. Minnesota 1973. An econometric analysis of state cigarette taxes, prices, and demand, with estimates of tax-induced interstate bootlegging.

JUDITH K. MANN, Ph.D. California (Los Angeles) 1974. The equilibrium of service firms.

ARTHUR H. MARTEL, Ph.D. Massachusetts 1974. An economic analysis of the market structure of the hotel-motel industry.

THOMAS G. MARX, Ph.D. Pennsylvania 1973. Diesel locomotive industry: A study in monopoly.

ALLEN K. MIEDEMA, Ph.D. North Carolina State 1974. Factor demands and particulate emission control regulations: The case of steam-electric power plants.

JOSEPH G. MOOSALLY, Ph.D. Iowa 1973. Risk aversion in the merger decision: An empirical study.

ROBERT E. NIELSEN, Ph.D. Iowa 1973. The structure and performance of selected, 4-digit textile mill product industries with special emphasis on the effects of imported textile mill products.

JAMES W. OWENS, Ph.D. North Carolina State 1973. Firm efficiency and market structure analysis of merger guidelines in the textile mill products industry.

ANNE E. PECK, Ph.D. Stanford 1973. Futures markets, supply response, and price stability.

DAVID PERRY, Ph.D. Pennsylvania 1974. Diversity and the demand for television broadcasting.

JAMES PIERSON, Ph.D. New York 1974. The advertising behavior of the regulated public utility.

GLENWORTH A. RAMSEY, Ph.D. Boston College 1974. Excessive advertising in market structures characterized by collusive pricing and conflict.

ALISON L. REASER, Ph.D. California (Los Angeles) 1974. Advertising and profitability.

KENNETH T. ROSEN, Ph.D. Massachusetts Institute of Technology 1974. Seasonal fluctuations in residential construction.

RICHARD K. RUDEL, Ph.D. Colorado State 1973. Economics of the ammonia industry.

MARIO M. SALINAS, Ph.D. Louisiana State 1974. A statistical study of cost and its implications for the efficient provision and pricing of electricity in the United States.

BRION R. SASAKI, Ph.D. Cincinnati 1974. A regional model of the future demand for transportation: The case of barge transportation.

CHARLES R. SCHERER, Ph.D. Cornell 1973. Estimating returns to scale in thermal electric power systems.

MICHAEL S. SHERMAN, Ph.D. Pittsburgh 1974. An investigation of the economic behavior and performance of private electric utilities in the United States.

JOHN W. SMYTHE, Ph.D. Nebraska 1973. The structure, conduct, and performance of the precious gem industry.

ROBERT M. SPANN, Ph.D. North Carolina State 1973. The supply of natural resources: The case of oil and natural gas.

AUSTIN SPENCER, Ph.D. Indiana 1973. An examination of relative downward industrial price flexibility, 1870–1921.

RODNEY E. STEVENSON, Ph.D. Michigan State 1973. Postal pricing problems and production functions.

ROBERT J. STONEBRAKER, Ph.D. Princeton 1973. The effect of risk on corporate rates of return.

CHANG LIH TUNG, Ph.D. Houston 1973. An analysis of cross-section production functions at 4-digit SIC manufacturing industries.

JERALD H. UDINSKY, Ph.D. California (Berkeley) 1974. The economic and legal aspects of stock market trading prior to the first major public announcement of corporate merger.

EDMON D. VILLANI, Ph.D. Pennsylvania 1973. Determinants of development costs in the chemical industry: An econometric investigation.

PATRICK J. WELCH, Ph.D. Pittsburgh 1973. An analysis of the effects on competition of patent licenses which restrict licensees with respect to fields of use and which discriminate between licensees with respect to royalty payments.

JAMES A. ZELLNER, Ph.D. Virginia Polytechnic Institute 1974. The behavior of regulating commissions: A case study of the Virginia State Corporation Commission.

Agriculture and Natural Resources

BAMIDELE ABOGUNRIN, Ph.D. Iowa State 1974. Educational and occupational plans and realizations of Iowa farm boys.

BAKIR ABED ABUKISHK, Ph.D. Iowa State 1973. Efficient utilization of resources within a regional framework.

TILAK ANAND, Ph.D. New York 1974. The food problem in India: Experience under India's Five-Year Plans.

DANIEL J. AROSTEGUY, Ph.D. Colorado State 1974. Projections of wildlife related recreation to 1985.

SJARIFUDDIN BAHARSJAH, Ph.D. North Carolina State 1974. The domestic and international trade of Indonesian coconut products.

TULIO BARBOSA, Ph.D. Purdue 1973. A normative analysis of land reform measures in the priority area of Rio de Janeiro, Brazil.

CHRISTOPHER D. S. BARTLETT, Ph.D. Stanford 1974. Factors affecting the spread of rain-grown cotton in tropical Africa.

BALA N. BATAVIA, Ph.D. North Carolina State 1974. Specification errors and empirical estimation of generalized crop-fertilizer production functions using cross-section data.

VINAY K. BHARGAVA, Ph.D. Illinois 1974. Effect of publicly supported credit programs on economic growth of small farmers in District Budaun, India.

T. ROY BOGLE, Ph.D. Missouri 1974. Farm investment analysis: Hand-calculated comparative budgeting and computerized five-year transitional budgeting.

JON A. BRESLAW, Ph.D. California (Berkeley) 1974. Economic analysis of water pollution in San Francisco Bay.

METIN CAGLAR, Ph.D. Iowa State 1974. Optimization of intraseasonal water allocation.

GIORGIO CANARELLA, Ph.D. Virginia 1973. Optimal policies for depletable resources.

RICHARD W. CARKNER, Ph.D. Michigan State 1974. A case study of the economic impacts of farm soil loss controls.

ROBERTO J. CASTRO, Ph.D. North Carolina State 1973. Effects of tariffs on prices, production, and trade on winter green peppers.

DENNIS M. CONLEY, Ph.D. Iowa State 1973. An analysis of the domestic and foreign distribution of *U.S.* heavy grains under alternative assumptions regarding possible future supply, demand, and transport costs.

JOHN T. CUMMING, Ph.D. Tufts 1973. Supply response in peasant agriculture: price and nonprice factors.

JOSE P. R. DE CASTRO, Ph.D. Purdue 1974. An economic model for establishing priorities for agricultural research and a test for the Brazilian economy.

EVONIR B. DE OLIVEIRA, Ph.D. Purdue 1974. Agricultural employment and land reform in Brazil.

OTTO C. DOERING III, Ph.D. Cornell 1973. Malaysian rice policy and the Muda River irrigation project.

PAUL DYGERT, Ph.D. Michigan 1973. Estimation of the cost of aircraft noise to residential activities.

WILLIAM A. FISCHER, D.B.A. George Washington 1974. Imposed R&D: An examination of the effects of environmental control requirements on industrial R&D programs.

JOHN A. GINZEL, Ph.D. Purdue 1973. Computerized budgetary projection of farm plans adaptable to farm accounting systems.

RICHARD H. GOLDMAN, Ph.D. Stanford 1974. Seasonal production instability and price formation in Indonesian rice markets, 1951–70.

PETER N. GORHAM, Ph.D. Southern Illinois 1974. A critique on traditional two sector models: The green revolution in India.

JOHN D. GRAHAM, Ph.D. Purdue 1973. A multiperiod cash management analysis of problems facing a commercial meat packing firm.

WARREN K. GREELEY, Ph.D. Tufts 1974. An analysis of the efficiency and equity of pollution control on the Androscoggin River.

RICHARD GREENHALGH, Ph.D. Missouri 1974. Identification of the Missouri public's perception of natural resource problems.

GREGORY C. GUSTAFSON, Ph.D. California (Berkeley) 1973. The California Land Conservation Act of 1965: Economic analysis of a new tool of land use policy.

MARK R. GUSTAFSON, Ph.D. California (Berkeley) 1974. An analysis of selected economic effects of rural subdivision development activity upon the public and private sectors of Tuolumne County, California, 1970.

ROBERT M. HARMEL, Ph.D. Iowa State 1974. Analysis of seven decisions to locate industry in rural areas of Iowa.

KENNETH J. HOCK, Ph.D. Arizona 1973. Agricultural adjustments to a falling groundwater table in central Arizona.

PETER N. HOPCRAFT, Ph.D. Stanford 1974. Human resources and technical skills in agricultural development: An economic evaluation of educative investments in Kenya's small farm sector.

URI HOROWITZ, Ph.D. Iowa State 1974. A dynamic model integrating demand and supply relationships for agricultural water, applied to determining optimal intertemporal allocation of water in a regional water project.

WASFY B. ISKANDER, Ph.D. Indiana 1973. The economic potential of copper in the Arctic.

GEOFFREY H. JACKSON, Ph.D. Cornell 1973. Milk supply response and some regional implications for dairy policy in the United States.

LARRY D. JONES, Ph.D. Purdue 1973. An experimental business gaming approach for examining behavior and performance in the food retailing industry.

S. JAMALADIN KALANTAR, Ph.D. Arizona 1973. Interregional price flexibilities and structural changes in the *U.S.* fed beef industry.

NATALAJA KALONJI, Ph.D. Missouri 1973. Decision making process of African farmers: A theoretical approach.

GEORGE E. KENNEDY, Ph.D. Purdue 1973. A simulation analysis of alternative schemes to stabilize hog-pork prices.

RODNEY C. KITE, Ph.D. Purdue 1973. An interregional analysis of livestock use of selected feed ingredients.

JEFFREY KOLB, Ph.D. Oregon 1974. An econometric study of discoveries of natural gas and oil reserves in the United States, 1948–70.

WON WHE KOO, Ph.D. Iowa State 1974. Linear programming models applied to interregional competition of grain transportation and production.

BUN SONG LEE, Ph.D. Southern Methodist 1974. Quality of inputs, returns to scale and the form of the production function in the Japanese tuna longline fishery.

CHAE WOONG LEE, Ph.D. Southern Methodist 1974. Production and externalities in the Japanese tuna fishery.

ALLAN E. LINES, Ph.D. Purdue 1973. A computerized model for planning the growth and organization of swine.

RALPH LUKEN, Ph.D. Michigan 1974. Efficiency and distributive consequences of preserving San Francisco Bay wetlands.

DON MCCLATCHY, Ph.D. Purdue 1973. Combines in the corn belt: A normative study of farm machinery demand.

THOMAS L. MANN, Ph.D. Iowa State 1973. Economic impact of restricting feed additives in livestock and poultry production.

JOHN F. MARTEN, Ph.D. Purdue 1973. An economic analysis of the future of the liquid feed supplement industry in Indiana.

STEPHEN F. MATTHEWS, Ph.D. Missouri 1974. The limited partnership as an investment vehicle in agriculture.

PHILLIP MAXWELL, Ph.D. Georgia 1973. The economic performance of small watershed development projects.

DALE J. MENKHAUS, Ph.D. Purdue 1973. The effects of environmental legislation on the structure of the pesticide industry: A simulation study.

PAUL W. MERKENS, Ph.D. Cornell 1974. Risk acceptance in public water supply: A New York City case study.

PETER R. MOOCK, Ph.D. Columbia 1973. Managerial ability in small-farm production: An analysis of maize yields in the Vihiga division of Kenya.

SUDHIN K. MUKHOPADHYAY, Ph.D. Minnesota 1974. Sources of variation in agricultural productivity: A cross-section time-series study of India.

PAUL H. MUTSCHLER, Ph.D. Pittsburgh 1973. Three suggested methods of controlling air pollution and their impact on the coal industry.

PETER N. NEMETZ, Ph.D. Harvard 1973. The economics of water pollution control in the British Columbia metal-mining industry.

KENNETH J. NICOL, Ph.D. Iowa State 1974. A modeling approach to the economic and regional impacts of sediment loss control.

DELBERT C. OGDEN, Ph.D. Colorado 1974. An economic analysis of air pollution: An emission-rights approach.

WAYNE L. OLSON, Ph.D. Missouri 1973. A review and reappraisal of speculative limit establishment procedures as applied to the Chicago Board of Trade wheat, corn, and soybean markets under noncrisis conditions.

SAVAS OZATALAY, Ph.D. Northwestern 1974. Farm supply functions: A decision theory context.

LALIT K. PATI, Ph.D. Missouri 1974. Development of experimental information in agricultural response research.

NEIL A. PATRICK, Ph.D. Iowa State 1974. Development and application of a model for multicounty rural community development.

J. B. Penn, Ph.D. Purdue 1973. Econometric policy models of commodity supply response.

Thomas R. Pierson, Ph.D. Cornell 1974. Economies of size in grocery distribution centers.

Ronald R. Piggott, Ph.D. Cornell 1973. A farm level econometric model of the *U.S.* apple industry: Applications to forecasting and analyzing policy issues.

Robert A. Richardson, Ph.D. Purdue 1973. An analysis of the generation of household solid waste residuals.

Abdullah A. Saleh, Ph.D. Purdue 1974. An analysis of consumer food expenditures and consumption waste.

Ali B. Z. Salem, Ph.D. Cornell 1973. Firm size and changes in industrial structure: An econometric analysis of the dairy industry in New York State.

William H. Schaffer, Jr., Ph.D. Cornell 1974. An economic analysis of nitrogen, phosphorus, and soil loss from agricultural production affecting water quality in a small New York watershed.

Herbert D. Schellenberg, Ph.D. Iowa State 1973. Identification, measurement, and incorporation of environmental quality objectives in natural resource development.

Michael G. G. Schluter, Ph.D. Cornell 1973. The interaction of credit and uncertainty in determining resource allocation and incomes on small farms, Surat District, India.

Wilson E. Schmeisseur, Ph.D. Purdue 1973. Optimal entrepreneurial decisions and investments of expanding corn belt dairy farmers at alternative milk prices, interest rates, and entrepreneurial abilities.

Gerald D. Schwab, Ph.D. Purdue 1974. A computerized decision-making model for the beef/forage enterprise.

Jerry Sesco, Ph.D. Southern Illinois 1974. Economics of pollution and the impact of environmental concern on the primary forest products industry: An integral synthesis.

John D. Shillingford, Ph.D. Cornell 1974. Financial potential and welfare implications of sugar cane harvest mechanization on Jamaican plantations.

Lawrence N. Smith, Ph.D. Purdue 1973. Remote sensing requirements, impact, and benefits in the soybean market and in the economy.

John J. Soladay, Ph.D. Northwestern 1974. The theory and measurement of natural resource depletion.

Craig E. Stanley, Ph.D. Claremont School 1974. The estimation of activity demand for a nonpriced recreational resource: Mineral King, California.

Arthur Stoecker, Ph.D. Iowa State 1974. A quadratic programming model of United States agriculture in 1980: Theory and application.

Jay L. Strom, Ph.D. Purdue 1973. Simulation of a swine breeding herd.

Werner J. Sublette, Ph.D. Arizona 1974. Demand for and value of outdoor recreation in the Salt-Verde Basin of Arizona.

Wallen Summers, Ph.D. Harvard 1973. The externalities of paper and paper board.

Veraphol Suvannunt, Ph.D. Iowa State 1973. Measurement of quantities and prices of product qualities.

Bock T. Tan, Ph.D. North Carolina State 1973. Prices and trade prospects for Malaysian palm oil.

Tesfai Tecle, Ph.D. Cornell 1973. An economic evaluation of agricultural package programs in Ethiopia.

J. V. Venkataram, Ph.D. Illinois 1974. Financing farm innovations in Mandya District, Mysore, India.

Melvin E. Walker, Jr. Ph.D. Illinois 1974. An economic evaluation of the impact of the commercial nitrogen control at the farm level.

Milton C. Weinstein, Ph.D. Harvard 1973. Bringing the outside inside: Government decision making for the environment.

Charles V. White, Ph.D. Ohio State 1973. Supply response and interregional competition in the Midwest dairy industry: An econometric analysis.

Wayne R. Willey, Ph.D. California (Berkeley) 1974. The diffusion of pest management information technology.

Carlos A. Zulberti, Ph.D. Cornell 1974. The feasibility of beef cattle feedlots in Argentina: Methodology of economic evaluations.

Manpower, Labor, and Population; including Trade Unions and Collective Bargaining

Joseph R. Antos, Ph.D. Rochester 1974. Supply and demand factors in the market for public school teachers.

J. Gregory Ballentine, Ph.D. Rice 1974. The role of residence choice in analysis of migration.

LEE S. BALLIET, Ph.D. Texas (Austin) 1974. Anglo poverty in the rural South.

CLARENCE H. BARNES, Ph.D. Tennessee (Knoxville) 1973. An investigation of intrafirm upward mobility among selected skilled craftsmen.

BURT S. BARNOW, Ph.D. Wisconsin (Madison) 1973. The effects of Head Start and socioeconomic status on cognitive development of disadvantaged children.

ROBIN L. BARTLETT, Ph.D. Michigan State 1974. The impact of married women and related factors on the distribution of family earnings.

MAURICE B. BETTER, Ph.D. Wisconsin (Madison) 1973. The collective bargaining process: Case study and analysis.

HOWARD C. BIRNBAUM, Ph.D. Harvard 1974. Determinants of earnings: The economic effect of job experience and the effect of cohorts on age-earnings profiles.

FARRELL E. BLOCH, Ph.D. Stanford 1973. The allocation of time to market and nonmarket work within a family unit.

BRYAN L. BOULIER, Ph.D. Princeton 1974. Two essays in the economics of dentistry: A production function for dental services and an examination of the effects of licensure.

MURRAY BROWN, Ph.D. Chicago 1974. Experience and earnings of male physicians in the United States.

MONCEF M. CHEIKHROUHOU, Ph.D. California (Berkeley) 1973. On-the-job training programs: A microeconomic analysis.

KIMON J. CONSTAS, Ph.D. State University of New York (Binghamton) 1974. Factor substitution and complementarity in Greek manufacturing, 1961–70.

ROBERT L. CROSSLIN, Ph.D. Missouri 1973. Asking wage behavior of unemployed workers.

JOHN P. DANFORTH, Ph.D. Northwestern 1974. Individual labor market behavior under uncertainty.

MARK DANIEL, Ph.D. Minnesota 1974. Problems in the theory of labor supply.

ROBERT C. DAUFFENBACH, Ph.D. Illinois (Urbana) 1973. The structure of occupational mobility in the *U.S.* economy.

DONALD L. DAVIS, Ph.D. Utah 1972. A conceptual manpower model of the life cycle of human resource.

SERGIO DECASTRO, Ph.D. Chicago 1973. Differences in the Santiago labor market.

LUCIA F. DUNN, Ph.D. California (Berkeley) 1974. Labor supply in southern industrialization.

ROBERT N. EBERT, Ph.D. Case Western 1974. Labor markets for selected Protestant clergy.

CHARLES J. ELLARD, Ph.D. Houston 1974. An investigation of the influence of the in-school neighborhood youth corps on earning capacity in Houston, Texas.

DENNIS H. ERICKSON, Ph.D. North Carolina 1973. An investigation into the determinants of migration.

PHILIP ERNST, Ph.D. New York 1973. The appropriate unit for collective bargaining under federal labor legislation.

GEORGE J. EUSKIRCHEN, Ph.D. Cincinnati 1974. Operating expenditures, enrollment, and economies of size: An economic analysis of higher education.

ALAN A. FISHER, Ph.D. California (Berkeley) 1974. The problem of teenage unemployment.

LAWRENCE FRIC, Ph.D. Toronto 1973. The role of commercial employment agencies in the Canadian labor market.

DAVID GAHTAN, Ph.D. New York 1974. Education and production: A manpower production model.

HERMAN F. GALLASCH, JR., Ph.D. North Carolina State 1973. The impact of minimum wage legislation on the farm labor market.

ROBERT M. GAY, Ph.D. California (Los Angeles) 1973. On-the-job training costs and their determinants in military occupations.

CURTIS GILROY, Ph.D. State University of New York (Binghamton) 1974. Investment in human capital and the nonwhite-white unemployment differential.

GERALD P. GLYDE, Ph.D. Illinois (Urbana) 1973. Economic analysis of employee turnover rates.

KENNETH GORDON, Ph.D. Chicago 1973. Accident rates and wages on *U.S.* class-I railroads.

PETER H. GREENWOOD, Ph.D. Brown 1974. The construction and testing of an alternative model of racial discrimination in the labor market.

ALFREDO GUTIERREZ, Ph.D. Michigan 1974. Economic determinants of Spanish intra-European migration 1963–71.

ABRAHAM E. HASPEL, Ph.D. Pennsylvania 1974. Occupational decision making: A socioeconomic analysis.

WALLACE E. HENDRICKS, Ph.D. California (Berkeley) 1974. Factors affecting the level of union wages in manufacturing.

KAREN C. HOLDEN, Ph.D. Pennsylvania 1973. A comparative study of labor force change: Japan 1920–69 and United States 1890–1955.

DAVID E. HORLACHER, Ph.D. Pennsylvania State 1973. Econometric estimation of the benefits of family planning in India.

DESMOND B. HUGHES, Ph.D. Princeton 1974. Overtime and fringe barriers to employment: Automobile assembly.

FRANK L. HULL, Ph.D. Illinois (Urbana) 1973. Economic effects of unionization of migrant labor.

TAKENORI INOKI, Ph.D. Massachusetts Institute of Technology 1974. Aspects of German peasant emigration to the United States 1815–1914: A reexamination of some behavioral hypotheses in migration theory.

MICHAEL C. KEELEY, Ph.D. Chicago 1974. A model of marital formation: The determinants of the optimal age at first marriage.

ROBIN KENDRICK, Ph.D. Princeton 1973. Regional markets for seasonal hired farm labor in the United States.

HWANG JOE KIM, Ph.D. Massachusetts 1973. An analysis of the interindustry wage structure of mining and manufacturing industries in Korea.

JONATHAN KING, Ph.D. California (Los Angeles) 1973. Social inequality and labor force participation.

EUGENE KIRK, Ph.D. Boston College 1974. A theoretical and empirical study of the impact of collective negotiations on public school teachers' salaries in the Commonwealth of Massachusetts.

PRADEEP KUMAR, Ph.D. Queen's 1974. Relative wage differentials in Canadian industries.

ALFRED LUBELL, Ph.D. New York 1974. Salary differentials among teachers of economics in the State University of New York.

KATHLEEN MCELROY, Ph.D. Case Western 1974. Labor relocation assistance: The Alabama experience.

BARBARA D. MCNEILL, Ph.D. Columbia 1973. A critique of uses of the rational model in educational planning and decision making.

ELIZABETH A. MILNE, Ph.D. Northwestern 1974. Attributes of the labor force productivity, and relative wages.

WESLEY N. MUSSER, Ph.D. California (Berkeley) 1974. Federal manpower policy and the rural sector.

DUANE A. OLSEN, Ph.D. Missouri 1973. The immigrant response to the industrialization of four distressed regions.

LUCIAN W. PALMER, Ph.D. Houston 1973. An examination of the rising levels of educational enrollments in Latin America.

JOSE PEREZ, Ph.D. Michigan 1973. The rate of return to educational investment with special reference to Puerto Rico.

ULKU POLAT, Ph.D. New York 1974. The role of dependency rate in economic development and growth, and income distribution: A theoretical approach.

JOHN POWEL, JR., Ph.D. Washington 1973. A theory of union behavior applied to the medical profession.

WILLIAM V. RICE, Ph.D. Louisiana State 1974. An inquiry into the evolving federal labor relations system with emphasis on private sector comparisons and contrasts.

STEPHEN ROACH, Ph.D. New York 1973. Short-run dynamics and the *U.S.* labor market: A costs-of-adjustment approach.

RICHARD D. ROBBINS, Ph.D. North Carolina State 1973. Costs and returns to occupational education.

HARVEY S. ROSEN, Ph.D. Harvard 1974. The impact of *U.S.* tax laws on the labor supply of married women.

STEPHEN W. SALANT, Ph.D. Pennsylvania 1973. Job search unemployment: The role of aggregate demand and heterogeneous supply.

STEVEN H. SANDELL, Ph.D. Minnesota 1973. Male-female salary differences among scientists with Ph.D.s.

DONALD T. SANT, Ph.D. California (Berkeley) 1974. Voluntary employment: An analysis of search and turnover.

PAWAN SAWHNEY, Ph.D. Boston 1974. Interindustry wage structure in Indian manufacturing: 1953–65.

WILLIAM E. SCHRANK, Ph.D. Wisconsin (Madison) 1973. Canadian job search—labor turnover relations: An empirical study in the use of monthly data.

JACKIE F. SHANER, Ph.D. North Carolina State 1973. Subsidization of public higher education in North Carolina.

KENNETH H. SHAPIRO, Ph.D. Stanford 1974. Efficiency

and modernization in African agriculture: A case study in Geita District, Tanzania.

LOIS B. SHAW, Ph.D. Michigan 1973. Metropolitan labor markets and labor migration.

VICTOR J. SHEIFER, Ph.D. New York 1974. Technological change and collective bargaining: Experiences in the New York milk-distributing industry, 1939–72.

STEPHEN SLAVIN, Ph.D. New York 1973. An evaluation of the economic cost and effectiveness of the Barbados family planning association.

SHARON P. SMITH, Ph.D. Rutgers 1974. Wage differentials between federal government and private sector workers.

VERNON K. SMITH, Ph.D. Michigan State 1973. The employment and earnings of AFDC mothers: The first-year effect of the earnings exemption in two Michigan counties.

ERIC J. SOLBERG, Ph.D. Claremont School 1974. Labor supply and labor force participation decisions of the AFDC population-at-risk.

EDWARD STEINBERG, Ph.D. New York 1973. Intrafirm and intraindustry mobility patterns and their implications for government-sponsored upgrading programs.

TOSHIAKI TACHIBANAKI, Ph.D. Johns Hopkins 1973. Quality change in labor input and wage differentials: A study of Japanese manufacturing industries.

MICHAEL B. TANNEN, Ph.D. Brown 1974. The distribution of labor income over time.

RICHARD H. THALER, Ph.D. Rochester 1974. The value of saving a life: A market estimate.

ALLEN R. THOMPSON, Ph.D. Texas (Austin) 1973. Comparative occupational position of white and nonwhite females in the United States.

BOONE TURCHI, Ph.D. Michigan 1973. The demand for children: An economic analysis of fertility in the United States.

JAMES D. VAN ERDEN, Ph.D. Utah. An analysis of social criteria and triggering mechanisms for extended unemployment benefits.

GEORGE M. VREDEVELD, Ph.D. Indiana 1973. Income redistributive effects of subsidizing public higher education in Indiana.

JAMES L. WALKER, Ph.D. Texas (Austin) 1974. Economic development, black employment, and black migration in the nonmetropolitan deep South.

ROBERT I. WEINSTEIN, Ph.D. Texas (Austin) 1974. Native Americans in rural and urban poverty.

RANDALL WEISS, Ph.D. Harvard 1974. Sources of change in the occupational structures of American manufacturing industries, 1950–60: An application of production function analysis.

HANDY WILLIAMSON, JR., Ph.D. Missouri 1974. Job turnover among displaced agricultural workers within the Mississippi Delta area of eastern Arkansas: Extent, costs, and criteria of cause.

IRA WINAKUR, Ph.D. American 1973. The income determinants of the Gallaudet College alumni.

TOM S. WITT, Ph.D. Washington (St. Louis) 1974. The regional racial unemployment rate differential.

PETER W. WYETH, Ph.D. California (Berkeley) 1974. An economic evaluation of trade union power in California agriculture.

JARED J. YOUNG, Ph.D. Southern California 1974. Discrimination, income, human capital investment, and Asian-Americans.

STEVEN P. ZELL, Ph.D. Massachusetts Institute of Technology 1974. A comparative study of the labor market characteristics of return migrants and nonmigrants in Puerto Rico.

Welfare Programs; Consumer Economics; Urban and Regional Economics

FRED ABRAHAM, Ph.D. Oregon 1973. An analysis of the redistributive effects on a one-and-one-half percent income tax for Lane County, Oregon.

ANDREAS A. ANDRIKOPOULOS, Ph.D. Southern California 1973. The determinants of growth differentials and regional concentration: A theoretical and empirical investigation.

ELIZABETH J. ARNAULT, Ph.D. Pennsylvania 1973. Profit maximizing expenditures for maintenance of New York City's rent-controlled housing.

KAZEM ATTARAN, Ph.D. Southern California 1973. Inter-metropolitan area budget cost determinants: An econometric analysis.

ROBERT C. BAESEMANN, Ph.D. California (Santa Barbara) 1973. The dynamics of agglomeration.

GARY T. BARNES, Ph.D. North Carolina State 1973. An econometric analysis of the college going and college choice decision.

HAROLF C. BARNETT, Ph.D. Massachusetts Institute of Technology 1974. The heroin epidemic: An analysis of the spread of heroin addiction in the United States.

MAX D. BENNETT, Ph.D. Johns Hopkins 1973. Health insurance and price distortions.

CHRISTINE E. BISHOP, Ph.D. Harvard 1974. Hospital wage structures in the Boston labor market.

CHARLES W. BLOWERS, Ph.D. Tennessee (Knoxville) 1973. Intermetropolitan reallocation of manufacturing activity: A theoretical and empirical analysis.

TONY BOPP, Ph.D. Missouri (Columbia) 1974. The economics of prisons.

SUSAN R. BRESLER, Ph.D. Georgia State 1974. Multiplier for a public program in a poverty area: Case study of the comprehensive health services program in Lowndes County, Alabama.

ERIC S. BROWN, Ph.D. Wisconsin (Madison) 1973. The effects of income expectations on housing purchase and consumption decisions.

RICHARD R. CARROLL, Ph.D. Kentucky 1973. The structure of employment in the area development districts of Kentucky, 1960–70: Analysis and implications.

CHARLES T. CLOTFELTER, Ph.D. Harvard 1974. An economic analysis of the effect of school desegregation on residential location and private school enrollment.

JOHN P. COMBS, Ph.D. North Carolina State 1973. The economics of information, discrimination in the real estate market, and the open housing law of 1968.

PHILIP J. COOK, Ph.D. California (Berkeley) 1974. The effect of legitimate opportunities on the probability of parolee recidivism.

GARY D. COOPER, Ph.D. Florida State 1974. The economics of ghetto expansion: A study of residential segregation.

PAUL N. COURANT, Ph.D. Princeton 1973. Economic aspects of racial prejudice in urban housing markets.

MELVIN L. DAVIS, Ph.D. Colorado State 1973. Rural development in Colorado.

DAVID L. DEBERTIN, Ph.D. Purdue 1973. An econometric investigation of the provision for public education in Indiana.

VERNON J. DIXON, Ph.D. Princeton 1973. A determination of investment priorities in urban black communities: Bedford-Stuyvesant.

THOMAS P. DRINKA, Ph.D. Iowa State 1973. A theoretical model of urban roadway pricing, with application to Dallas, Texas.

JOSEPH E. EARLEY, Ph.D. Fordham 1974. Empirical test of an urban land-use model.

JAMES H. EBLEN, Ph.D. Tennessee (Knoxville) 1974. A study of outmigration of workers from southeastern metropolitan areas.

ULRICH ERNST, Ph.D. Indiana 1973. Human resources in regional economic growth.

WILLIAM A. FISCHEL, Ph.D. Princeton 1973. Fiscal and environmental considerations in the location of firms in suburban communities.

ROBERT F. FIX, Ph.D. Iowa State 1973. Financing elementary and secondary education in Iowa: An economic analysis and empirical results.

MARC GAUDRY, Ph.D. Princeton 1973. The demand for public transit in Montreal and its implications for transportation planning and cost-benefit analysis.

ROBERT F. GILLINGHAM, Ph.D. Pennsylvania 1973. Place to place rent comparisons using hedonic quality adjustment techniques.

WILLIAM J. GOGGIN, Ph.D. Iowa State 1973. The measurement of well-being in need analysis models.

JOHN GOODMAN, Ph.D. Michigan 1974. Local residential mobility and family housing adjustments.

WILLIAM GOOLSBY, Ph.D. Wisconsin (Milwaukee) 1974. Marginal cost pricing of city services.

PAUL L. GRIMALDI, Ph.D. Fordham 1974. An empirical analysis of the noninstitutional health services provided by the New Jersey health services program during calendar years 1970 and 1971.

GERALD GUNN, Ph.D. Colorado State 1973. Economic aspects of desegregation.

RICHARD L. HANEY, JR., D.B.A. Indiana 1974. Some determinants of the value of housing services.

TIMOTHY H. HANNAN, Ph.D. Wisconsin (Madison) 1974. The economics of methadone maintenance.

DAVID HEMENWAY, Ph.D. Harvard 1974. Industry-wide voluntary product standards.

MARK S. HENRY, Ph.D. Kansas State 1974. Projecting state tax revenues from a regional input-output model: Kansas taxes, 1970–80.

THOMAS W. HIESTAND, Ph.D. Kansas State 1974. Simulating the impact of import substitution on a regional economy.

LEO HILLMAN, Ph.D. Pennsylvania 1973. Collective consumption, optimum resident size, and locational choice.

RONALD K. HOFER, Ph.D. Indiana 1973. The economic benefits of medicare to the aged: A micro-simulation study.

WILLIAM L. HOLAHAN, Ph.D. Brown 1974. Hospital behavior in a spatial economy.

STANLEY KEIL, Ph.D. Oregon 1973. A cost-benefit study of the Oregon and Washington alcohol safety action projects.

TIMOTHY KERSTEN, Ph.D. Oregon 1973. Public higher education in Oregon: Who gets it? Who pays for it?

RICHARD H. KLEM, Ph.D. Northwestern 1974. Investment in moveable equipment by hospitals.

NORMAN L. KNAUB, Ph.D. Pennsylvania State 1974. The effect of welfare payments on family expenditures.

WILLIAM R. LATHAM III, Ph.D. Illinois (Urbana) 1973. Agglomerative economies as a factor in the location of manufacturing industry.

AARON LEVINE, Ph.D. New York 1973. Economic analysis of educational vouchers.

JAMES E. LONG, Ph.D. Florida State 1974. Racial discrimination in employment: A public-private sectoral analysis.

GERALD S. MCDOUGALL, Ph.D. Claremont School 1974. The household choice problem in a spatial context.

JAMES B. MCFADYEN, Ph.D. Kentucky 1973. Postwar southern manufacturing: Concentrations and factor intensities.

WILLIAM K. MACREYNOLDS, Ph.D. Southern California 1974. Contributions to the theory of industrial location: Urban decentralization.

ARNOLD R. MILLER, Ph.D. Nebraska 1974. Spatial variations in the Nebraska economy.

TAKAHIRO MIYAO, Ph.D. Massachusetts Institute of Technology 1974. Dynamics and optimality in the theory of urban location.

JAMES P. MOODY, Ph.D. California (Berkeley) 1974. Testing tax capitalization: An experiment afforded by a local public transit improvement.

MARILYN L. MOON, Ph.D. Wisconsin (Madison) 1974. The economic welfare of the aged: A measure of economic status and an analysis of federal programs.

MARK H. MOORE, Ph.D. Harvard 1973. Policy toward heroin use in New York City.

F. PATRICIA MUNCH, Ph.D. Chicago 1973. An economic analysis of eminent domain.

MICHAEL P. MURRAY, Ph.D. Iowa State 1974. An econometric analysis of benefits to tenants in alternative federal housing programs.

FRANK J. NAVRATIL, Ph.D. Notre Dame 1974. The effect of relative school quality and other determinants on the demand for Catholic education: An econometric analysis of the archdiocese of St. Louis.

DAVID T. NELSON, Ph.D. Wisconsin (Madison) 1974. The effects of student and instructor characteristics on learning in an introductory economics course.

ANTHONY L. OSTROSKY, Ph.D. Pittsburgh 1973. The effective demand for health services: An empirical examination based on demographic analysis.

DUANE B. OYEN, Ph.D. Iowa 1973. Estimation of benefits and analysis of factors influencing land use change in the Iowa River flood plain.

THOMAS D. PEARSON, Ph.D. Georgia State 1974. Factors which influence the location of manufacturing in the nonmetropolitan United States.

HENRY O. POLLAKOWSKI, Ph.D. California (Berkeley) 1974. The effects of local public services on residential location decisions: An empirical study for the San Francisco Bay area.

IRWIN PRICE, Ph.D. Boston 1974. The social cost of airport noise as measured by rental changes: The case of Logan Airport.

DAVID L. PURYEAR, Ph.D. Princeton 1973. Central place theory and interurban specialization.

ALHAMBRA RACHMAN, Ph.D. Montana State 1974. An econometric analysis of the demand for beef in Japan.

DAVID REES, Ph.D. Oregon 1973. Economic aspects of punishment forms and practices.

MICHAEL REICH, Ph.D. Harvard 1974. Racial discrimination and the white income distribution.

CLIFFORD E. REID, Ph.D. Princeton 1973. The integration of urban spatial location and the economic theory of the allocation of time.

WILLIAM M. RHODES, Ph.D. Minnesota 1974. An economic analysis of the criminal courts.

ARTHUR J. ROBSON, Ph.D. Massachusetts Institute of Technology 1974. Pollution, congestion, and urban structure.

MARK E. RODEKOHR, Ph.D. Colorado 1974. The effects of declining growth rates on resource allocation in American educational systems.

DEVINDER K. SAGGAR, Ph.D. Cincinnati 1974. Residential location and housing consumption in the city of Cincinnati.

ROBERT SCHAFER, Ph.D. Harvard 1973. The suburbanization of multifamily housing.

BARTON A. SMITH, Ph.D. Houston 1974. The supply of urban housing.

NEAL M. SOSS, Ph.D. Princeton 1974. The demand for college education: The applications decision.

GARY S. STACEY, Ph.D. Ohio State 1973. The allocation of resources in municipal fire protection.

ROBERT F. STAUFFER, Ph.D. Virginia Polytechnic Institute 1974. Monetary trends in the United States: A regional analysis.

THOMAS F. STINSON, Ph.D. Minnesota 1973. The demand for education: An hedonic approach.

WALTER B. THOMPSON, Ph.D. Georgia 1974. Public policy toward substandard housing occupancy: Economic policy analysis.

W. TOM TRULOVE, Ph.D. Oregon 1973. Economics of paternalism: Federal policy and the Klamath Indians.

AYDIN ULUSAN, Ph.D. Michigan State 1973. An econometric analysis of Michigan's welfare caseload, 1968–71.

JAMES F. VEATCH, Ph.D. Illinois 1973. Mass transportation in small and medium size urban areas.

DAVID R. WHEELER, Ph.D. Massachusetts Institute of Technology 1974. Economic control and urban growth.

MICHELLE J. WHITE, Ph.D. Princeton 1973. A study of suburban fiscal zoning in fragmented metropolitan areas.

BARBARA L. WOLFE, Ph.D. Pennsylvania 1973. A socioeconomic analysis of family building behavior.

SALLY WRIGHT, Ph.D. Southern Illinois 1973. The political economy of the Economic Opportunity Act, 1964 to 1972.

JONG KEUN YOU, Ph.D. State University of New York (Binghamton) 1974. The demand for nonfarm housing: A time-series analysis.

JANET S. YOUNG, Ph.D. Pennsylvania 1973. Family size, income, and housing expenditure in the inner city rental market.

ANNOUNCEMENT

NOTICE TO ALL GRADUATE DEPARTMENTS

The December 1975 issue of the *Review* will carry the seventy-second list of doctoral dissertations in political economy in American universities and colleges. The list will specify doctoral degrees conferred during the academic year terminating June 1975. This announcement is an invitation to send us information for the preparation of the list. This announcement supersedes and replaces a letter which was sent annually from the managing editor's office.

The *Review* will publish in its December 1975 issue the names of those who will have been awarded the doctoral degree since June 1974 and the titles of their dissertations. Dissertation abstracts will no longer be published, as these are published elsewhere.

By June 30, please send us this information on 3×5 cards, conforming to the style shown below, one card for each individual. Please indicate by a classification number in the right-hand corner the field in which the thesis should be classified. The classification system is that used by the *Journal of Economic Literature* and printed in every issue.

JEL Classification No. ______

Name: LAST NAME IN CAPS: First Name, Initial

Institution Granting Degree: ______

Degree Conferred (Ph.D. or D.B.A.) ______ Year ______

Dissertation Title: ______

When degrees in economics are awarded under different names, such as Business Administration, Public Administration, or Industrial Relations, candidates in these fields whose training has been *primarily in economics* should be included.

All items and information should be sent to the Assistant Editor, *American Economic Review*, Box Q, Brown University, Providence, Rhode Island 02912.

THE AMERICAN ECONOMIC REVIEW

VOLUME LXIV

THE AMERICAN ECONOMIC ASSOCIATICN

Executive Office: Nashville, Tennessee

Editorial Office: Brown University, Providence, Rhode Island

CONTENTS OF ARTICLES AND SHORTER PAPERS

K. J. Arrow: Limited Knowledge and Economic Analysis 1

R. H. Day, S. Morley, and K. R. Smith: Myopic Optimizing and Rules of Thumb in a Micro-Model of Industrial Growth 11

Y. Onitsuka: International Capital Movements and the Patterns of Economic Growth 24

R. W. Parks: The Demand and Supply of Durable Goods and Durability 37

B. T. McCallum: Competitive Price Adjustments: An Empirical Study 56

R. G. Kuller and R. G. Cummings: An Economic Model of Production and Investment for Petroleum Reservoirs 66

J. Conlisk: Can Equalization of Opportunity Reduce Social Mobility? 80

S. K. Ghosh: Toward a Theory of Multiple Customs Unions 91

R. C. Vogel: The Dynamics of Inflation in Latin America, 1950–1969 102

R. Dusansky and P. J. Kalman: The Foundations of Money Illusion in a Neoclassical Micro-Monetary Model 115

S. Star: Accounting for the Growth of Output 123

S. J. Turnovsky: The Stability Properties of Optimal Economic Policies 136

J. P. Brown: Product Liability: The Case of an Asset with Random Life 149

M. Shane: Capital Markets and the Dynamics of Growth 162

M. J. Hartley and N. S. Revankar: Labor Supply Under Uncertainty and the Rate of Unemployment 170

D. P. Baron: Default Risk, Homemade Leverage, and the Modigliani-Miller Theorem 176

M. I. Kamien and N. L. Schwartz: Patent Life and R&D Rivalry 183

J. Heckman: Life Cycle Consumption and Labor Supply: An Explanation of the Relationship Between Income and Consumption Over the Life Cycle 188

G. D. Craig: Money Illusion and the Aggregate Consumption Function: Note 195

R. B. Porter: Semivariance and Stochastic Dominance: A Comparison 200

E. Baltensperger: The Precautionary Demand for Reserves 205

R. Bowden: Risk Premiums and the Life Cycle Hypothesis 211

L. Figà-Talamanca: Private and Social Rates of Return to Education of Academicians: Note 217

J. A. Tomaske: Private and Social Rates of Return to Education of Academicians: Note 220

J. F. McDonald: Housing Market Discrimination, Homeownership, and Savings Behavior: Comment 225

J. F. Kain and J. M. Quigley: Reply 230

J. F. Weston, S. Lustgarten, and N. Grottke: The Administered-Price Thesis Denied: Note 232

J. P. Mattila: Job Quitting and Frictional Unemployment 235

T. K. Kumar and E. Asher: Soviet Postwar Economic Growth and Capital-Labor Substitution: Comment 240

M. Weitzman: Reply 243

K. J. Arrow: General Economic Equilibrium: Purpose, Analytic Techniques, Collective Choice 253

J. J. Mingo: Capital Importation and Sectoral Development: A Model Applied to Postwar Puerto Rico 273

A. O. Krueger: The Political Economy of the Rent-Seeking Society 291

E. F. Fama: The Empirical Relationships Between the Dividend and Investment Decisions of Firms 304

G. V. G. Stevens: On the Impact of Uncertainty on the Value and Investment of the Neoclassical Firm 319

W. J. Stull: Land Use and Zoning in an Urban Economy 337

L. Gallaway and V. Shukla: The Neoclassical Production Function 348

D. Wisecarver: The Social Costs of Input-Market Distortions 359

J. M. Marshall: Private Incentives and Public Information 373

E. R. Berndt and L. R. Christensen: Testing for the Existence of a Consistent Aggregate Index of Labor Inputs 391

J. P. Gould and C. R. Nelson: The Stochastic Structure of the Velocity of Money 405

N. M. Gordon, T. E. Morton, and I. C. Braden: Faculty Salaries: Is There Discrimination by Sex, Race, and Discipline? 419

K. Borch: The Rationale of the Mean-Standard Deviation Analysis: Comment 428

G. O. Bierwag: Comment 431

H. Levy: Comment 434

S. C. Tsiang: Reply and Errata for Original Article 442

F. T. Dolbear, Jr.: Externalities as Commodities: Comment 451
T. Page and J. Ferejohn: Comment 454
R. A. Meyer: Reply 460
T. H. Tietenberg: On Taxation and the Control of Externalities: Comment 462
E. A. Thompson and R. Batchelder: Comment 467
W. J. Baumol: Reply 472
Y-H. Yeh: Consumer Preferences and Gains from Trade 473
L. B. Smith: A Note on the Price Adjustment Mechanism for Rental Housing 478
M. L. Wachter: Phase II, Cost-Push Inflation, and Relative Wages 482
B. F. Kiker and W. P. Liles: Earnings, Employment, and Racial Discrimination: Additional Evidence 492
J. N. Bhagwati and T. N. Srinivasan: On Reanalyzing the Harris-Todaro Model: Policy Rankings in the Case of Sector-Specific Sticky Wages 502
H. I. Grossman: The Nature of Quantities in Market Disequilibrium 509
E. C. H. Veendorp: The Nature of Quantities in Market Disequilibrium 515
A. O. Krueger: The 1974 Report of the President's Council of Economic Advisers: International Issues 523
D. Laidler: The 1974 Report of the President's Council of Economic Advisers: The Control of Inflation and the Future of the International Monetary System 535
F. Modigliani: The 1974 Report of the President's Council of Economic Advisers: A Critique of Past and Prospective Policies 544
W. D. Nordhaus: The 1974 Report of the President's Council of Economic Advisers: Energy in the Economic Report 558
P. A. Samuelson and S. Swamy: Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis 566
R. N. Batra: Technological Change in the Soviet Collective Farm 594
T. Johnson and F. J. Hebein: Investment in Human Capital and Growth in Personal Income 1956–1966 604
J. A. Hanson: Optimal International Borrowing and Lending 616
A. Takayama: On Biased Technological Progress 631
J. V. Henderson: The Sizes and Types of Cities 640
G. W. Douglas and J. C. Miller III: Quality Competition, Industry Equilibrium, and Efficiency in the Price-Constrained Airline Market 657
S. Rosefielde: Factor Proportions and Economic Rationality in Soviet International Trade 1955–1968 670
R. E. Schuler: The Interaction Between Local Government and Urban Residential Location 682
A. Razin: A Note on the Elasticity of Derived Demand Under Decreasing Returns 697
A. Sandmo: A Note on the Structure of Optimal Taxation 701
E. K. Browning: The Diagrammatic Analysis of Multiple Consumption Externalities 707
A. P. Lerner: Priorities and Pollution: Comment 715
J. L. Stein: Reply 718
R. Fels and C. E. Hinshaw: An Analysis of Turning Point Forecasts: A Fairly Polite Comment 724
H. O. Stekler: A Polite Reply 728
W. G. Waters II: Transport Costs and the Static Welfare Costs of Tariffs 730
E. Silberberg: The Theory of the Firm in "Long-Run" Equilibrium 734
S. S. Sidhu: Relative Efficiency in Wheat Production in the Indian Punjab 742
H. M. Hochman and J. D. Rodgers: Redistribution and the Pareto Criterion 752
J. E. Anderson: A Note on Welfare Surpluses and Gains from Trade in General Equilibrium 758
W. Schulze and R. C. d'Arge: The Coase Proposition, Information Constraints, and Long-Run Equilibrium 763
J. R. Marsden, D. E. Pingry, and A. Whinston: The Process Analysis Alternative to Statistical Cost Functions: Comment 773
J. M. Griffin: Reply 777
M. L. Greenhut, M. J. Hwang, and H. Ohta: Price Discrimination by Regulated Motor Carriers: Comment 780
J. E. Olson: Reply 785
W. Ramm: On the Durability of Capital Goods Under Imperfect Market Conditions 787
M. B. Krauss: Commodity Trade and Factor Mobility 797
D. Salkever: Choice Involving Unwanted Risky Events and Optimal Insurance: Comment 802
J. M. Parkin and S. Y. Wu: Reply 804
H. Hori: The Number of Firms and Competition: Comment 805
E. F. Fama and A. B. Laffer: Reply 806
W. Leontief: Structure of the World Economy 823

E. J. Kane: All for the Best: The Federal Reserve Board's 60th Annual Report......... 835

J. E. Stiglitz: On the Irrelevance of Corporate Financial Policy......................... 851

J. L. Stein: Unemployment, Inflation, and Monetarism............................ 867

G. E. Johnson and F. P. Stafford: The Earnings and Promotion of Women Faculty 888

S. S. Shalit and U. Ben-Zion: The Expected Impact of the Wage-Price Freeze on Relative Shares................................ 904

G. W. Scully: Pay and Performance in Major League Baseball......................... 915

B. Klein: Competitive Interest Payments on Bank Deposits and the Long-Run Demand for Money............................ 931

D. Levhari and Y. Weiss: The Effect of Risk on the Investment in Human Capital......... 950

H. P. Binswanger: The Measurement of Technical Change Biases with Many Factors of Production...................... 964

J. Muellbauer: Household Production Theory, Quality, and the "Hedonic Technique" ... 977

A. Feldman and A. Kirman: Fairness and Envy................................ 995

C. Henry: Investment Decisions Under Uncertainty: The "Irreversibity Effect"......... 1006

L. J. White: Industrial Organization and International Trade: Some Theoretical Considerations.............................. 1013

R. Cummings and V. Norton: The Economics of Environmental Preservation: Comment . 1021

A. E. Abrassart and D. D. McFarlane: Comment.................................. 1025

A. C. Fisher, J. V. Krutilla, and C. J. Cicchetti: Reply.................................. 1030

R. N. Batra and W. R. Russell: Gains from Trade Under Uncertainty................ 1040

G. C. Rausser: Technological Change, Production, and Investment in Natural Resource Industries.............................. 1049

D. M. G. Newbery: Cropsharing Tenancy in Agriculture: Comment................... 1060

P. K. Bardhan and T. N. Srinivasan: Rejoinder 1067

C. C. Blaydon and P. W. Marshall: Incentive Contracts and Competitive Budding: Comment.................................. 1070

D. P. Baron: Reply........................ 1072

T. Russell: Feige and Parkin on the Optimal Quantity of Money...................... 1074

Y. Niho: Population Growth, Agricultural Capital, and the Development of a Dual Economy.............................. 1077

P. L. Cheng: Comment on Corporate Income Taxes and the Cost of Capital: A Correction 1086

N. Minabe: Capital and Technology Movements and Economic Welfare............. 1088

CONTENTS OF PAPERS AND PROCEEDINGS

Richard T. Ely Lecture

R. M. Solow: The Economics of Resources or the Resources of Economics 1

The World Dynamics of Economic Growth

H. E. Daly: The Economics of the Steady State 15

W. D. Nordhaus: Resources as a Constraint on Growth 22

R. F. Kosobud and W. D. O'Neill: A Growth Model with Population Endogenous 27

Population and Economic Policy

R. G. Ridker: Resource and Amenity Implications of Population Changes 33

A. C. Kelley: The Role of Population in Models of Economic Growth 39

A. Sweezy and A. Owens: The Impact of Population Growth on Employment 45

International Energy Supply

J. W. McKie: The Political Economy of World Petroleum 51

M. A. Adelman: Politics, Economics, and World Oil 58

Lessons From and For the Current Economic Expansion

P. W. McCracken: Three Lessons 68

W. J. Fellner: A Strange Expansion 71

P. A. Samuelson: Lessons from the Current Economic Expansion 75

O. Eckstein: A Plethora of Lessons from the Recent Expansion 78

Two Years of Wage-Price Controls

R. F. Lanzillotti and B. Roberts: The Legacy of Phase II Price Controls 82

D. J. B. Mitchell and A. R. Weber: Wages and the Pay Board 88

B. Bosworth: The Inflation Problem during Phase III 93

M. T. Hamilton: Price Controls in 1973: Strategies and Problems 100

Monetary Policy and Sectoral Credit Flows

J. S. Duesenberry: Alternatives to Monetary Policy 105

A. F. Brimmer: Alternative Monetary-Fiscal Policies and Sectoral Credit Flows 112

Major Economic Problems of the 1970's

R. L. Heilbroner: The Clouded Crystal Ball 121

K. Gordon: Some Conjectures on Policy Problems of the 1970's 125

S. Bowles: Economists as Servants of Power. 129

G. H. Moore: Some Secular Changes in Business Cycles 133

H. S. Houthakker: Policy Issues in the International Economy of the 1970's 138

Concepts of Distributional Equity

J. Rawls: Some Reasons for the Maximin Criterion 141

D. C. Mueller: Achieving the Just Polity 147

J. M. Buchanan and W. C. Bush: Political Constraints on Contractual Redistribution 153

Discussion by **A. K. Klevorick** 158

New Dimensions of Economic Inequality

J. D. Smith and S. D. Franklin: The Concentration of Personal Wealth, 1922–1969 162

B. A. Okner and J. A. Pechman: Who Paid the Taxes in 1966? 168

B. L. Friedman and L. J. Hausman: Income Conditioning in a System of Transfer Programs 175

The Great Society Versus the New Federalism

C. L. Schultze: Sorting Out the Social Grant Programs: An Economist's Criteria 181

L. C. Thurow: Cash Versus In-Kind Transfers 190

The Fight Against Poverty

I. Garfinkel and R. Haveman: Earnings Capacity and the Target Efficiency of Alternative Transfer Programs 196

R. I. Lerman and A. A. Townsend: Conflicting Objectives in Income Maintenance Programs 205

The Property Tax: Progressive or Regressive?

H. Aaron: A New View of Property Tax Incidence 212

R. A. Musgrave: Is a Property Tax on Housing Regressive? 222

Discussion by **H. E. Brazer, D. Netzer, A. F. Friedlaender, E. R. Rolph, and G. E. Peterson** 230

Women at Work: At Home and In the Market

V. R. Fuchs: Recent Trends and Long-Run Prospects for Female Earnings 236

A. Leibowitz: Education and Home Production 243

M. J. Boskin: The Effects of Government Expenditures and Taxes on Female Labor 251

Production and the Aggregate Demand Function for Labor

T. J. Sargent and N. Wallace: The Elasticity of Substitution and Cyclical Behavior of

Productivity, Wages, and Labor's Share... 257
S. Rosen and M. I. Nadiri: A Disequilibrium Model of Demand for Factors of Production 264

The Political Economy of Multinational Enterprises

G. C. Hufbauer: Multinational Corporations and the International Adjustment Process 271
R. Vernon: Competition Policy toward Multinational Corporations.................... 276
G. H. Hildebrand: Problems and Policies Affecting Labor's Interests................. 283

Trends, Prospects and Policies for East-West Trade

G. Grossman: Prospects and Policy for U.S.-Soviet Trade........................... 289
A. Eckstein and B. Reynolds: Sino-American Trade Prospects and Policy............... 294
A. Brown, P. Marer, and E. Neuberger: Prospects for U.S.-East European Trade....... 300

Perspectives on Twentieth Century Economic Theory

J. Hicks: Capital Controversies: Ancient and Modern................................ 307

Is Economic Theory With It?

G. S. Becker: On the Relevance of the New Economics of the Family................. 317
V. L. Smith: Economic Theory and Its Discontents.................................. 320
H. G. Johnson: The State of Theory.......... 323
T. C. Koopmans: Is the Theory of Competitive Equilibrium With It?.................... 325

The Theory of Policy

T. Muench and N. Wallace: On Stabilization Policy: Goals and Models................ 330
W. P. Heller and K. Shell: On Optimal Taxation with Costly Administration........... 338

Allocating Resources for Policy Research

A. M. Rivlin: How Can Experiments Be More Useful?................................ 346
F. T. Juster: The Use of Surveys for Policy Research.............................. 355

Intercity Passenger Transportation

G. Eads: Airline Capacity Limitation Controls: Public Vice or Public Virtue?........ 365
E. P. Patton: Amtrak in Perspective: Where Goest the Pointless Arrow?............... 372
H. S. Norton: The Wheel: Should We Reinvent It?................................ 378

The Economics of the First Amendment

R. H. Coase: The Market for Goods and the Market for Ideas......................... 384
W. F. Baxter: Regulation and Diversity in Communications Media.................... 392
Discussion by **B. M. Owen**................. 400

Economic Education

R. Fels: Developing Independent Problem-Solving Ability in Elementary Economics... 403
B. H. Booms and D. L. Kaltreider: Computer-Aided Instruction for Large Elementary Courses............................... 408
K. Boulding and E. Boulding: Introducing Freshmen to the Social System............ 414

Endowment Income, Capital Gains and Inflation Accounting

D. A. Nichols: The Investment Income Formula of the American Economic Association 420
J. Tobin: What is Permanent Endowment Income?................................ 427
J. M. Litvack, B. G. Malkiel, and R. E. Quandt: A Plan for the Definition of Endowment Income........................ 433
Discussion by **R. Eisner and S. W. Black**..... 438

CONTRIBUTORS TO ARTICLES AND SHORTER PAPERS.

Abrassart, A. E. 1025
Anderson, J. E. 758
Arrow, K. J. 1, 253
Asher, E. 240
Baltensperger, E. 205
Bardhan, P. K. 1067
Baron, D. P. 176, 1072
Batchelder, R. 467
Batra, R. N. 594, 1040
Baumol, W. J. 472
Ben-Zion, U. 904
Berndt, E. R. 391
Bhagwati, J. N. 502
Bierwag, G. O. 431
Binswanger, H. P. 964
Blaydon, C. C. 1070
Borch, K. 428
Bowden, R. 211
Braden, I. C. 419
Brown, J. P. 149
Browning, E. K. 707
Cheng, P. L. 1086
Christensen, L. R. 391
Cicchetti, C. J. 1030
Conlisk, J. 80
Craig, G. D. 195
Cummings, R. G. 66, 1021
d'Arge, R. 763
Day, R. H. 11
Dolbear, F. T. Jr. 451
Douglas, G. W. 657
Dusansky, R. 115
Fama, E. F. 304, 806
Feldman, A. 995
Fels, R. 724
Ferejohn, J. 454
Figà-Talamanca, L. 217
Fisher, A. C. 1030
Gallaway, L. 348
Ghosh, S. K. 91
Gordon, N. M. 419
Gould, J. P. 405
Greenhut, M. L. 780
Griffin, J. M. 777
Grossman, H. I. 509
Grottke, N. 232
Hanson, J. A. 616
Hartley, M. J. 170
Hebein, F. J. 604
Heckman, J. 188
Henderson, J. V. 640
Henry, C. 1006
Hinshaw, C. E. 724
Hochman, H. M. 752
Hori, H. 805
Hwang, M. J. 780
Johnson, G. E. 888
Johnson, T. 604
Kain, J. F. 230
Kalman, P. J. 115
Kamien, M. I. 183
Kane, E. J. 835
Kiker, B. F. 492
Kirman, A. 995
Klein, B. 931
Krauss, M. B. 797
Krueger, A. O. 291, 523
Krutilla, J. V. 1030
Kuller, R. G. 66
Kumar, T. K. 240
Laffer, A. B. 806
Laidler, D. 535
Leontief, W. 823
Lerner, A. P. 715
Levhari, D. 950
Levy, H. 434
Liles, W. P. 492
Lustgarten, S. 232
McCallum, B. T. 56
McDonald, J. F. 225
McFarlane, D. D. 1025
Marsden, J. R. 773
Marshall, J. M. 373
Marshall, P. W. 1070
Mattila, J. P. 235
Meyer, R. A. 460
Miller, J. C. III, 657
Minabe, N. 1088
Mingo, J. J. 273
Modigliani, F. 544
Morley, S. 11
Morton, T. E. 419
Muellbauer, J. 977
Nelson, C. R. 405
Newbery, D. M. G. 1060
Niho, Y. 1077
Nordhaus, W. D. 558
Norton, V. 1021
Ohta, H. 780
Olson, J. E. 785
Onitsuka, Y. 24
Page, T. 454
Parkin, J. M. 804
Parks, R. W. 37
Pingry, D. E. 773
Porter, R. B. 200
Quigley, J. M. 230
Ramm, W. 787
Rausser, G. C. 1049
Razin, A. 697
Revankar, N. S. 170
Rodgers, J. D. 752
Rosefielde, S. 670
Russell, T. 1074
Russell, W. R. 1040
Salkever, D. 802
Samuelson, P. A. 566
Sandmo, A. 701
Schuler, R. E. 682
Schulze, W. 763

Schwartz, N. L. 183
Scully, G. W. 915
Shalit, S. S. 904
Shane, M. 162
Shukla, V. 348
Sidhu, S. S. 742
Silberberg, E. 734
Smith, K. R. 11
Smith, L. B. 478
Srinivasan, T. N. 502, 1067
Stafford, F. P. 888
Star, S. 123
Stein, J. L. 718, 867
Stekler, H. O. 728
Stevens, G. V. G. 319
Stiglitz, J. E. 851
Stull, W. J. 337
Swamy, S. 566
Takayama, A. 631
Thompson, E. A. 467
Tietenberg, T. H. 462
Tomaske, J. A. 220
Tsiang, S. C. 442
Turnovsky, S. J. 136
Veendorp, E. C. H. 515
Vogel, R. C. 102
Wachter, M. L. 482
Waters, W. G. II, 730
Weiss, Y. 950
Weitzman, M. 243
Weston, J. F. 232
Whinston, A. 773
White, L. J. 1013
Wisecarver, D. 359
Wu, S. Y. 804
Yeh, Y-H. 473

CONTRIBUTORS TO PAPERS AND PROCEEDINGS

Aaron, H. 212
Adelman, M. A. 58
Baxter, W. F. 392
Becker, G. S. 317
Black, S. W. 441
Booms, B. H. 408
Boskin, M. J. 251
Bosworth, B. 93
Boulding, E. 414
Boulding, K. 414
Bowles, S. 129
Brazer, H. E. 230
Brimmer, A. F. 112
Brown, A. A. 300
Buchanan, J. M. 153
Bush, W. C. 153
Coase, R. H. 384
Daly, H. E. 15
Duesenberry, J. S. 105
Eads, G. 365
Eckstein, A. 294
Eckstein, O. 78
Eisner, R. 438
Fellner, W. J. 71
Fels, R. 403
Franklin, S. D. 162
Friedlaender, A. F. 232
Friedman, B. L. 175
Fuchs, V. R. 236
Garfinkel, I. 196
Gordon, K. 125
Grossman, G. 289
Hamilton, M. T. 100
Hausman, L. J. 175
Haveman, R. 196
Heilbroner, R. L. 121
Heller, W. P. 338
Hicks, J. 307
Hildebrand, G. H. 283
Houthakker, H. S. 138
Hufbauer, G. C. 271
Johnson, H. G. 323
Juster, F. T. 355
Kaltreider, D. L. 408
Kelley, A. C. 39
Klevorick, A. K. 158
Koopmans, T. C. 325
Kosobud, R. F. 27
Lanzillotti, R. F. 82
Leibowitz, A. 243
Lerman, R. I. 205
Litvack, J. M. 433
Malkiel, B. G. 433
Marer, P. 300
McCracken, P. W. 68
McKie, J. W. 51
Mitchell, D. J. B. 88
Moore, G. H. 133
Mueller, D. C. 147
Muench, T. 330
Musgrave, R. A. 222
Nadiri, M. I. 264
Netzer, D. 231
Neuberger, E. 300
Nichols, D. A. 420
Nordhaus, W. D. 22
Norton, H. S. 378
Okner, B. A. 168
O'Neill, W. D. 27
Owen, B. M. 400
Owens, A. 45
Patton, E. P. 372

Pechman, J. A. 168
Peterson, G. E. 234
Quandt, R. E. 433
Rawls, J. 141
Reynolds, B. 294
Ridker, R. G. 33
Rivlin, A. M. 346
Roberts, B. 82
Rolph, E. R. 233
Rosen, S. 264
Samuelson, P. A. 75
Sargent, T. J. 257
Schultze, C. L. 181
Shell, K. 338
Smith, J. D. 162
Smith, V. L. 320
Solow, R. M. 1
Sweezy, A. 45
Thurow, L. C. 190
Tobin, J. 427
Townsend, A. A. 205
Vernon, R. 276
Wallace, N. 257, 330
Weber, A. R. 88

EMPLOYMENT SERVICES

NATIONAL REGISTRY FOR ECONOMISTS

The National Registry for Economists was established in January, 1966, to provide a centralized nationwide clearinghouse for economists on a year-round basis. It is located in the Chicago Professional Placement Office of the Illinois State Employment Service and is staffed by experienced placement personnel, operating under the guidance and direction of Regional and National Bureau of Employment Security Professional Placement officials, and in cooperation with the American Economic Association. It is a free service. The National Registry for Economists maintains completely separate listings from those of the American Economic Association, and the National Registry does *not* have the listings as shown in the *American Economic Review.* There are no registration, referral, or placement fees. Application and order forms used in the Registry are available upon request from the: National Registry for Economists, Professional Placement Center, 40 West Adams Street, Chicago, Illinois 60603.

AMERICAN ECONOMIC ASSOCIATION
VACANCIES AND APPLICATIONS
To Be Phased Out

In October the American Economic Association began publication of *Job Openings for Economists* (*JOE*), a bimonthly listing of job vacancies. Subscriptions to *JOE* are available to members of the Association, departments of economics, and other organizations at an introductory rate of $6 (USA, Canada, Mexico, first class) or $11 (foreign airmail) for the first year. For more information, write *JOE,* American Economic Association, 1313 21st Avenue South, Nashville, Tennessee 37212. Please note that only *jobs* will be listed.

The Vacancies and Applications section of the *American Economic Review* will no longer be published after this issue, which lists only applicants for positions.

Economists Available for Positions

Italics indicate fields of specialization

Computer Assisted Instruction: Economist with interest and experience in computer assisted instruction in economics desires part-time, full-time, or consulting work in this area. E2589

Economics: Man, 25, married. B.A., M.A., in economics. Excellent grades, references. Desires position in economics in New York area. Strong background in microeconomics and familiar with mathematical analysis. Résumé will be furnished upon request. E2594

International Economics, Economic Development, Microeconomics, Input-Output Economics, Labor Economics, Cost-effectiveness Studies, Economics of Natural Resources: Man, 46, Ph.D. 1963. Trilingual; several years of experience in teaching, government and business in Europe, Latin America, Canada, and U.S.A. Seeks teaching or research position. Ready to relocate. E2603

Economist-Administrator: Ph.D. (economics-finance & management) and J.D. (law). Male, 44, Married, U.S. nationality, business and academic (administration and faculty) experience, publications. American Men of Science; Who's Who in Consulting; Other. Currently, university administration. Desire administrative (Chairman, Dean, Vice-President, or President) or faculty professorship in applied economics-finance or management position. E2606

Urban & Regional Economics; Resource Economics; Land Use; Agricultural Economics: Man, 31, married, Ph.D. Three years experience in economic planning for central city in metropolitan area. Publications in economic base, housing market, and socio-demographic analyses and in capital budgeting. Served as projects leader. Desires position with corporation, federal or state government or quasi-public establishment. E2539

Quantitative Economic Analysis, Systems Analysis, Cost Analysis and Modeling, Program Evaluation, Cost-Benefit Analysis: Man, 44, M.S. plus additional graduate study. Twenty years of experience in research and analysis of social/economic problems. Experience includes teaching, program planning and evaluation, cost-benefit analysis, economic analysis of aerospace, defense, welfare, medical, judicial, education, manpower development, state and local government planning, and other programs. Currently manager of an economic and cost analysis group of professionals in a non-profit corporation. Desire challenging managerial/analysis assignment. E2516

Macro Theory and Policy, Human Capital and Labor: Man, Ph.D., 40, married, experienced teacher (including graduate), publications, extensive faculty governance and committee experience. Desire full professorship with administrative possibilities, preferably not in Boston-to-Washington corridor. Excellent references. Available September 1975. Résumé upon request. E2617

Public Finance, Economic Development, Economic History, Economic Systems, Economic Theory, Business Courses: Man, 44, married; B.A. Honors, M.A., D.Sc. Eleven years of teaching experience; currently associate professor of economics; publications. Seeks teaching or research position. Available in September 1974. E2618

Public Finance, Urban Economics, Micro and Macroeconomics: Man, 39, Ph.D. Considerable teaching, research and administrative experience; publications. Desires research or teaching position with opportunity for research in the area of public finance, particularly state and local finance. Available July 1975. E2619

Development, Urban and Housing Economics, Econometrics: 57, D.Sc., Econ. Sciences. Available for senior position in the USA, Canada, or with international organizations, for university teaching and/or research. Background of 15 years teaching experience and about 20 years of research experience in the Central Government, United Nations work in the Far-East and Caribbean. Special eminence in organizing international seminars and training courses in the field of development planning, housing and urban economics, and the like. E2620

Consumer Economics, Forecasting, Quantitative Economics, Economic Theory: Man, 37, married; Ph.D. in quantitative economics. Has nine years experience in teaching economics and statistics. Now tenured at a large midwest university. Desires to move to a more challenging position. Publications in top journals. Enjoys teaching at both undergraduate and graduate levels. Available September 1974. Résumé on request. E2621

Macro, Micro, Money and Banking, Public Finance, Introductory Statistics, International Economics, History of Economic Thought, Principles, Economic Systems, Agricultural Economics, Market Structure and Asian Economics: Man 43, married, one child, B.S., LL.B., M.S. & Ph.D., 1962 and 1965, from leading land grant universities. 10 years undergraduate teaching and three years research experience. Presently tenured full professor in a Midwestern state college. Excellent teacher with very good student evaluations. Modest publications, including a small book. Six years administrative experience. Wish to relocate; location open but prefers East, Mideast, Southeast, South and Midwest. Seeks College or University teaching and/or research position. Interested in undergraduate teaching, graduate program not essential. Would consider research position in private firms. References and résumé available on request. E2622

International Trade, Human Capital, Economic Development, Economic Theory: Man, 29, married, Ph.D., some publications. Teaching and research experience in the United States, Middle East, Africa. Consulting experience international organizations. Seeks teaching position West, Middle West, or Northeast. Available September 1975. Résumé on request. E2623

International Economics, Trade, International Finance, Economic Development: Man, 56, married, B.A. (Chicago), M.A. (Fletcher School of Law and Diplomacy), all work completed for Ph.D. except thesis. Has had almost 30 years of U.S. government experience, practically all with the Department of State and U.S. Foreign Service in the international economic field. Has held major supervisory and program direction positions such as Director of State Department's Office of International Trade and Finance; Minister for Economic and Commercial Affairs in American Embassy, Bonn; Minister for Political and Economic Affairs, American Embassy, New Delhi; Counselor for Economic Affairs, American Embassy, Yugoslavia; Deputy and Acting Director of U.S. Aid Mission, Yugoslavia; and Deputy Director of State Department's Research Bureau. Seeks senior executive position of operating or advisory character in international economic field with business or banking institution, private foundation or inter-governmental or other international agency. Available immediately. Résumé on request. E2624

Statistics (Including Sampling), Computer Applications Econometrics, Development Economics, National Accounts: Man, 30, married. B.A., M.A., Ph.D. (Economics). Five years teaching with emphasis on quantitative economics; four years as a statistician with the Statistical Office of the United Nations. Publications. Desires position in the New York Metropolitan Area. Résumé available on request. Available on reasonable notice. E2625

Economics, Finance, Cost Analysis: Female, 22, single. B.A. in economics with minors in political science and United States history. Plans to work on graduate degree in economics at night. Free to

travel or relocate. Some experience as tax research assistant. Seeks a position as an economist or a job related to finance or cost analysis. References available upon request. E2626

Economics and Business: Man, Ph.D. Will take a one-year assignment. Dr. Clyde Crobaugh, 3300 S. Ocean Blvd., Delray Beach, Florida 33444.

Economic Gerontology, Natural Resource, Environmental, and Energy Economics, Micro-Macro Economics, Business and Government, and Industrial Organization: Man, 40, married Ph.D., dissertation in environmental economics; one year as government economist; last four years as assistant professor at midwest liberal arts college, interdisciplinary grants experience; seeks challenging position in teaching, research, and/or administration. Available June 1975. Résumé on request. E2627

Economics-Business Administration: Male, 45, married, Ph.D. (International Finance), J.D., seeks position teaching some combination of business finance courses, business law courses, and international finance. Publications. Extensive teaching, administrative, law practice and business experience. E2628

Economics and Business Administration, Micro, Macro, Mathematical Economics, Econometrics and Statistics, Money and Banking, Public Finance, Managerial Finance, Managerial Economics, Quantitative Management, Economic Development, International Economics, Economic Systems: Ph.D. in economics, male, 41, 10 years of dynamic teaching experience and flexible at graduate and undergraduate levels. Currently an associate professor. Requests for résumé and interview are welcome. E2629

Microeconomic Theory, Labor Economics, Consumer Economics, History of Thought, Economic History: Man, 30, married; A.B., M.A., Ph.D. in dissertation stage. Seven years teaching and consulting experience; fellowships, scholarly papers and publications. Major text on consumer economics. Prefer upper midwest small college part or full time. E2630

Population/Family Planning (Economic Analysis, Surveys, Evaluation): Man, 38, married, M.A. (Economics); 7 years as junior economist with World Bank team and federal government in India; 3 years as statistician with international division of American organization in New York City. U.S. permanent resident. E2631

Principles, Micro and Macro Theory, Labor and Manpower Economics, Monetary Theory: Man, 36, married, Ph.D. Five years of teaching experience, three years as graduate teaching assistant, two years as assistant professor and as lecturer for the University of Maryland, European Division; administrative experience as administrative assistant to the president of a liberal arts college. Desires teaching position in U.S. or Canada. E2632

Finance, Accounting, Business Economics: Ph.D. (ABD), M.B.A., family man. Thirteen years diversified teaching in finance, accounting, investments, economics at large eastern schools; quantitative and computer oriented; 14 years broad industrial, financial and administrative experience; some publications. Impeccable Ivy credentials. Wish to consider relevant teaching and/or administrative post. E2633

Public Finance, International Economics and Finance, Economic Development, Money and Banking: Man, 49, Ph.D., Wisconsin. Experience includes teaching at the undergraduate and graduate level, staff economist with international organizations, consulting work here and abroad. Fluent in Spanish and German, and reading knowledge of several other languages. Some publications. U.S. citizen. Now teaching and consulting in Washington, DC. Seeks teaching, research or staff position for 1975. Will relocate for position expected to be permanent. E2634

International Trade and Finance, Industrial Organization, Economic Theory, Quantitative Economics, Principles: Man, 34, married. B.A., M.A., Ph.D. Six years teaching experience in liberal arts university, two years as department chairman. Currently economic policy fellow of Brookings Institution assigned to research position in federal agency. Publications. Desire teaching/administrative/research position. Résumé on request. Available September 1975. E2635

Business Conditions Analysis, Macroeconomics, Econometrics, Transportation Economics, Managerial Economics, Corporate Planning: Man, 30, married. Ph.D. from leading private university. National Defense Fellow. Woodrow Wilson Fellow. Presently manager of economic forecasting for U.S. corporation with annual sales in excess of one billion dollars. Responsibilities include management of a four-man staff of economists in the areas of business conditions analysis and forecasting. Prior teaching experience in college of business at large state university. Excellent teaching references. Seeks college/university teaching position. Résumé on request. E2636

Economic Theory, Economic History, Economic Development, Population: Man. Single. Ph.D. Eighteen years teaching experience at undergraduate and graduate levels. Fulbright Fellowship. Business experience. Desires teaching and/or research and/or administrative position. Résumé on request. E2637

Economist: Experienced economist, Cornell Ph.D., seeks teaching/research position starting Fall 1975, in basic macro/micro theory; labor economics; labor relations including collective bargaining, labor law, industrial relations theory, labor unions; manpower economics; and comparative international labor problems. Has several years' teaching experience and some publications. E2638

Micro and Macroeconomics, Business Economics, History of Economic Thought, Methodology, Organization Theory, Data Processing Management: Male, 29, married. M.S. Bus. Econ. Ph.D. candidate, dissertation in progress, degree expected summer 1975. Four years teaching experience. Three years CPT U.S. Army as data processing manager for U.S. Army Logistics Center with responsibility for computer simulations used in logistical war games. Desires teaching or research position at institution with large or growing computer facility. Résumé on request. E2639

Resource and Environmental Economics, Income Distribution Theory, General Economic Theory: Male, 28, Ph.D. expected Spring, 1975. B.S. physics, M.I.T. 1968. M.A. economics, University of Colorado, 1973. Desires full-time research position or academic position with possibility of teaching graduate theory courses. Will consider any unusual or particularly challenging positions in academia, private or public research oriented institution, government, etc. Especially interested in position utilizing prior scientific background. Foreign also considered. Rudimentary knowledge of Spanish, elementary fluency Russian. Wife, Ph.D. Environmental Biologist also seeks position. Available Fall 1975. Résumés on request. E2640

Industrial Organization, Government Regulation of Business, Microeconomics, Managerial Economics: Man, 33, married; Ph.D. Four years teaching experience at graduate and undergraduate level at a major state university. Good research and publications record. Seeks teaching or corporate research position. Prefers southwestern or southeastern U.S. Résumé available upon request. E2641

Quantitative Economic Analysis, Systems Analysis, Cost Analysis and Modeling, Program Evaluation, Cost-Benefit Analysis: Man, 44, M.S. plus additional graduate study. Twenty years of experience in research and analysis of social/economic problems. Experience includes teaching, program planning and evaluation, cost-benefit analysis, economic analysis of aerospace, defense, welfare, medical, judicial, education, manpower development, state and local government planning, and other programs. Currently manager of an economic and cost analysis group of professionals in a non-profit corporation. Desire challenging managerial/analysis assignment. E2516

Applied Econometrics and Statistics, Managerial Economics, Micro- and Macro-theory: Age 33, Ph.D., married. Eight years undergraduate and graduate teaching experience at well-known private university on the East coast. Research and publications in economics of natural resources and in development economics. Desires position on the faculty of a university or college which encourages close student-teacher contact and which has good computer facilities. Résumé, reprints, and references on request. E2642

Econometrician: Canadian citizen, specialised in spatial econometrics & urban planning with ten years experience and degrees from four countries (USA, Canada, Nederland and India) now currently finishing his Ph.D. doctoral dissertation here in the Netherlands School of Economics, Nederland, available for research work from Aug/Sept. 1975. Papers presented at the Econometric Society, & one scheduled for presentation for the International Federation of Automatic Control, 6th World Congress, Aug. 1975. Please Contact GKS/ C/o Dr. B. Pazderka, (School of Business, Queens, Kingston, Ont.) 2000 Front Road, Kingston, Ont. K7M 4M1.

Development Economics, International Economics, Economic Theory, Input-Output Economics, Latin American Economics, Econometrics: Man, 35, married, Ph.D. Copious teaching and consulting experience. Presently employed on faculty of a major state university. Extensive publications including four books (three of which were edited or co-edited) and numerous articles in leading scholarly journals. Active in professional associations. In short, all the trappings of a successful academic career. Seeks something more challenging and rewarding, possibly in government or business. Possesses languages, experience abroad, an interdisciplinary inclination, and leadership qualities. E2643

Economics and Business: Man, 32, married, Ph.D. Seven years teaching experience. Publications. Consulting. Former U.S. Army officer. Currently department chairman. Seeks similar or administrative appointment. Vita on request. E2644

Human Capital, Economics of Education, Economics of Discrimination, Public Finance, Urban Economics, Micro: Woman, 31, single; Minnesota Ph.D. Four years teaching experience, one year as policy analyst with HEW, two years at major academic research institute. Desires teaching/research position, preferably in western part of the country. Available September 1975. E2645

Public Finance, Public Expenditure Analysis, Welfare Economics, Government Budgeting, Benefit/Cost Analysis, Public Administration, Man, 37, Ph.D.; 8 years as federal government economist involved in both defense and social programs. Currently, also teaching B/C Analysis and government budgeting in graduate school of government and business administration of major eastern university. Desire full-time teaching position, preferably on West coast, with opportunity to work with local governments. Available June 1975. E2646

Regional and Urban Economics, Monetary, Development, Labor, Theory: Economist, man, 33, seeks position with government agency or private firm to serve as research economist. Ph.D., M.S., B.A. from three recognized schools. Six years full-time college teaching experience. Research experience, coauthor of several professional studies. Currently employed as professor. Available July 1975. E2580

Lipsey & Steiner

FOURTH EDITION

February 1975

Harper & Row, Publishers